Annotated Bibliography of
BLS Statistical Papers
1996-2000

Compiled by the
Office of Survey Methods Research
U.S. Bureau of Labor Statistics
2 Massachusetts Ave. N.E.
Washington DC 20212

January 2001

Note: The Annotated Bibliography is now available on the internet. In addition to the abstracts, many of the papers are also available. The location is:

http://stats.bls.gov/orehome.htm

(Authors who were BLS employees at the time of the writing of each paper are shown in boldface type.)

Baskin, Robert M. and **Leaver, Sylvia G.** (1996), "Estimating The Sampling Variance For Alternative Forms Of The U.S. Consumer Price Index," *Proceedings of the Section on Survey Research,* American Statistical Association, 192-197.

The current official U.S. Consumer Price Index is a Laspeyres index. An alternative form of the index known as a geometric mean index will be published in 1997. The sampling variance for the Laspeyres form of the index has been previously estimated and published. In this work the sample variance for the geometric mean index is studied. A simulation is performed to study properties of variance estimators. Initial estimates in the housing component for the geometric mean index and an alternative form of the Laspeyres index have already been estimated by the Bureau. Empirical work will compare the variance estimates of these alternative indices using linearization and vplx software.

Black, Susan R., Ernst, Lawrence R. and **Tehonica, Jason** (1997), "Sample Design And Estimation For The National Compensation Survey," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 457-462.

The National Compensation Survey (NCS) is the new BLS survey for measuring employee wages by skill level. This paper discusses the sample design, estimation and variance estimation procedures for this survey. The survey uses a rotating panel design, with three stages of selection used in selecting each sample panel, namely: geographic area PSUs; establishments selected from industry strata; and occupations selected separately from each sample establishment. The weighting incorporates several stages of adjustment, including three stages of nonresponse adjustment. Among the issues discussed are a new method for weighting the occupational selections which correctly adjusts for employment changes over time; a method for redistributing the weights of nonrespondent establishments to

respondent establishments that is different from the approach commonly used; and the allocation of the sample in each industry strata among noncertainty PSUs in a way that ensures that establishments in different PSUs with the same sampled employment have the same unconditional probability of selection.

Black, Susan R., Ernst, Lawrence R. and **Tehonica, Jason** (1997), "Statistical Problems In BLS Compensation Surveys When Collected Establishment Data Differs From The Assigned Data;" *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 463-469.

The National Compensation Survey (NCS) is the new BLS survey for measuring employee wages by skill level. The sample design for the NCS includes the selection of a sample of establishments and then the selection of a sample of occupations within each selected establishment. Special sampling and weighting procedures are required for these stages of selection when the data collected are not what was originally intended. The issues discussed include: collection from a unit that is other than the assigned unit, as for example when establishments merge; situations where, because of respondent or interviewer burden, subsampling of physical locations that comprise an assigned establishment or employees within an occupation are required; and a single central selection of occupations that is done for certain large companies instead of a separate selection for each sample establishment within the company.

Bosley, John, Dashen, Monica, and **Fox, Jean** (1999), "Effects on List Length and Recall Accuracy of Order of Asking Follow-up Questions About Lists of Items Recalled in Surveys," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Surveys often require respondents to recall a number of items belonging to a class or category, and also to recall significant facts or attributes about each recalled item. For example, the survey may ask for recall of purchases within a category such as "furniture" during some reference period, and to also recall for each item retrieved from memory some descriptive details (item attributes), such as price, date of purchase, model number, and the like. This purpose of the study was to examine the effects of two different ways of presenting a complex free recall task on the amount and accuracy of information retrieved from memory. The recall task focused on retrieving multiple instances of item within named categories, along with descriptive item details. The study findings are intended to apply to recall of purchased items, and so an experimental procedure simulating actual purchases was carried out on all study participants. Their simulated

purchase and pertinent details about each one were recorded to serve as a basis for assessing accuracy of recall. After an interval during which they worked on a distracter task, participants were presented with a recall task. In one condition, they were asked to recall several items within a category before being asked to supply detail about any of the items. In the second condition, they supplied details about each item as soon as it was recalled. Outcome measures included the number of items per category recalled, accuracy of recall of both items and item detail (attributes), ratings of perceived task difficulty, and frequency of verbal and non-verbal behaviors indicative of fatigue, boredom and other signs of motivational or affective reactions to the task, or of cognitive burden. Study findings will be presented and discussed from the perspective of what cognitive and motivational factors account for any differences in the effectiveness of the two methods for eliciting information of this nature.

Bradley, **Ralph** (1997), "Detecting Unobservable Inventory Accumulation," *Proceedings of the Section on Survey Research Methods* , American Statistical Association.

Often times when a retailer puts a nonperishable item on sale, consumers will "stock-up" on this item. The maintenance of household inventories play a role in consumer efforts to reduce their Cost of Living(COL). Data that is recorded from scanner sales in supermarkets cannot directly measure what part of a sale will be consumed immediately and what part will be stored. This paper derives simple models on how the consumers observes prices and then uses this information to store inventories. We then use the A.C. Nielsen Academic database for canned tuna fish sales to estimate an errors component model that dissagregates unit sales into current consumption and current inventories.

Kennickell, Arthur, B., Barrett, Don and **Bradley, Ralph** (1997), "Collection Of Survey Data On Sexual Orientation For Research Purposes," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Outside of HIV/AIDS-related work, a relatively small amount of information has been collected in major surveys on sexual attraction, identification, or behavior. Because homosexuals are a "rare" population, the number of observations on which such information is collected is often insufficient for hypothesis testing. It seems likely that nonsampling errors are also serious. Many people are uncomfortable revealing their sexual preferences, and for some people even the terms that are used to describe sexual orientation vary in important ways. This roundtable will focus on the needs for such information for research and the obstacles to collecing it.

Bradley, Ralph, (2000); "Finite Sample Effects in the Construction of Price Indexes," *Proceedings of the Section on Government Statistics,* American Statistical Association.

No abstract available.

Bradley, Ralph (1996), "The Use Of Scanner Data To Reduce The Mean Squared Error Of The CPI," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

There are several factors that contribute to the mean squared error of the CPI. We derive these sources and then determine how a change from the current collection system to a system that has access to almost the entire population of price quotes will reduce the mean squared error. Our analysis focuses on the bias reductions that should come from using true weights rather than imputed weights, and on deriving a sampling scheme that will minimize the variance of the BLS estimator.

Brown, Sharon (1997), "Potential Uses Of Continuous Measurement Estimates In BLS Labor Force Programs," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

The Bureau of Labor Statistics (BLS) is responsible for the measurement of labor force statistics for the nation, States and substate areas (some with very small populations). The Current Population Survey (CPS), conducted by the U.S. Bureau of the Census for BLS, provides accurate and timely monthly national statistics. For States and substates, the CPS sample sizes are too small to yield estimates with meaningful reliability. The Local Area Unemployment Statistics (LAUS) program, therefore, uses data from a variety of sources to develop models for labor force estimates for more than 6,700 subnational areas. The sources include the CPS, BLS's Current Employment Statistics (CES) survey, the unemployment insurance program and the decennial census. In this paper, we discuss potential uses of data from the continuous measurement program, based on the American Community Survey (ACS), to improve local and national labor force estimates. Possible enhancements include: 1) better substate population figures that can be used in preparing LAUS estimates; 2) address refinement for CPS sampling; 3) the use of superior population controls in national estimation and tighter monitoring of coverage ratios for demographic subgroups; 4) improved estimation of the component of CPS variance due to first-stage sampling of large geographic areas called Primary Sampling Units (PSUs); and 5) superior models for substate areas that utilize auxiliary data from ACS.

Butani, Shail J., and **McElroy, Michael** (1999), " *Managing Various Customer Needs for* Occupational Employment Statistics and Wage Survey," Proceedings of the Section on Survey Research Methods, American Statistical Association.

Originally, Occupational Employment Statistics (OES) Survey was designed to measure occupational employment by industry at the state level. For example, number of electrical and electronic engineers in research and testing services industry (Standard Industrial Classification Code 873). In the early 1990's, about 15 states asked guidance from U.S. Bureau of Labor Statistics to collect and produce wage rates for occupations. In response to the increasing demand for occupational wage rates, the OES sample expansion was implemented in all states plus District of Columbia with the 1996 survey. In this paper, we describe the major changes made to the survey design in order to meet the needs of new sponsors, states, and other data users. The implementation of the new design took place in a federal/state environment under an extremely tight time schedule

Butani, Shail, Robertson, Kenneth and **Mueller, Kirk** (1998), "Assigning Permanent Random Numbers to the Bureau of Labor Statistics Longitudinal (Universe) Data Base," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 451-462.

The U.S. Bureau of Labor Statistics (BLS) routinely administers several large ongoing establishment surveys. Many employers are asked to supply information for several of these BLS surveys each year. Recently, efforts have been made to spread the burden more evenly to employers by controlling the probability that they will be selected for multiple BLS surveys. This control is obtained by utilizing Permanent Random Numbers (PRNs). This paper describes the initial and ongoing assignment of PRNs to the BLS Longitudinal Database (LDB). This assignment is done in a manner that meets competing program requirements. The sample selection strata definitions and sample unit definitions are not consistent for the participating surveys, which provides some unique and interesting challenges for this project. Another major problem to overcome is how to ensure that new establishments are being properly represented when the samples for ongoing surveys are updated. We use collocated PRNs to solve this problem.

Butani, S., Stamas, G. and Brick, Michael (1997), "Sample Redesign For The Current Employment Statistics Survey," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 517-522

> In this paper, we present some issues as well as empirical results pertaining to the sample redesign for the Bureau of Labor Statistics' (BLS) Current Employment Statistics (CES) Survey, commonly known as the payroll survey. CES is a survey of 390,000 businesses that is conducted monthly in a Federal/State cooperative environment to produced estimates of employment, payroll, and hours worked. Because the month-to-month movements in the CES series are closely followed by policy makers and forecasters, the thoroughness of the research on many aspects of the sample redesign is extremely critical. In this study, we cover: 1) treatment of multi-establishment firms during sample allocation, selection and estimation phase; 2) procedures for stratification, atypical units, and imputation; 3) reliability of estimates at various geographical vs. industrial levels; 4) stratified random sampling vs. probability proportional to size designs; 5) the usage of permanent random numbers for frame maintenance, updating probabilities of selection, and sample rotation; and 6) performance of alternative estimators in presence of nonresponse with respect to levels, month-to-month change, revisions between preliminary and final monthly estimates, and March-to-March revisions as measured by a "known population" that lags in time and is based on administrative records. Multi-year simulations are performed for 11 large States using the data from BLS' Business Establishment List (BEL).

Butani, S., Harter, Rachel, and Wolter, Kirk (1997), "Estimation Procedures For The Bureau Of Labor Statistics Current Employment Statistics Program," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 523-528

> In this paper, authors summarize the research and recommendations with respect to the estimation process and identify known issues that remain to be research. They give a brief overview of the steps in estimation after the collection of sample data. The new estimation procedures include: (1) editing procedures for erroneous data, (2) adjustment for misaligned data, (3) adjustments for outliers, (4) imputation for missing data, (5) simple unbiased estimation, (6) estimation incorporating ES-202 employment data as an auxiliary variable, (7) estimation for births/deaths, (8) composite estimation, (9) seasonal adjustment, (10) variance estimation, and (11) benchmarkings or realignment to ES-202 employment data.

Butani, Shail, Cahoon, Lawrence and **Kojetin, Brian**, (1996) "Federal Government Shutdown: Options For CPS Data Collection" *Proceedings of the Section on Survey Research,* American Statistical Association 705-710.

The Federal Government shutdowns in November 1995 and again in December 1995 that extended into January 1996 created extraordinary challenges for data collection activities pertaining to the Current Population Survey (CPS). The CPS is a national labor force survey of over 50,000 households conducted monthly by the Bureau of Census for the Bureau of Labor Statistics (BLS). For November 1995 through January 1996 data, officials at both agencies considered options such as: not collecting for a particular month, changing the reference week, delaying collection, extending the collection period, and using accelerated (e.g., overtime) collection procedures. Changes in data collection procedures for CPS could, among other things, result in recall bias, different rotation group bias, and distort seasonal adjustment factors. We examined the effect of an increase in noninterviews resulting from closing down the survey prior to completion of interviewing. The research conducted on several months of prior data indicated that the characteristics of the late respondents differed from the early respondents. In particular, the expected values of some key labor force estimates were different.

Cano, Stephanie, **Getz, Patricia**, **Kropf, Jurgen, Scott, Stuart** and **Stamas, George** (1996), "Adjusting For A Calendar Effect In Employment Time Series" *Proceedings of the Section on Survey Research,* American Statistical Association, 656-661.

The Bureau of Labor Statistics Current Employment Statistics (CES) program is a monthly survey of nearly 400,000 business establishments nationwide; its reference period is the pay period including the 12th of each month. Of paramount importance to most CES data users are over-the-month changes in total nonfarm employment levels; thus, the seasonal adjustment of these data critically affects the analysis of national employment trends. This study investigates a calendar effect which may cause difficulties in interpreting movements in the current seasonally adjusted series. The effect arises because there are sometimes 4 and sometimes 5 weeks between the reference periods in any given pair of months (except February/March). This varying interval effect is estimated by an ARIMA model with regression variables, using the Bureau of the Census X-12-ARIMA procedure. Estimates of the interval effect are tested for significance. Series estimated with and without the interval adjustment are compared for smoothness and stability across successive time spans. Differences between over-the-month employment changes as

previously published and as estimated with interval effect modeling are examined.

Chun, Young I. (1997), **"**Nonresponse Follow-up in Establishment Surveys: A Split-half Experiment," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 988-991..

Efforts to convert nonrespondents in establishment surveys require understanding the compliance process specific to the organizational environment, and employing follow-up methods from this theoretical consideration. A nonresponse follow-up approach evaluated in this paper is developed from a theory of compliance in establishment surveys. This approach is subsequently tested against a technique currently employed in the Hours at Work Survey conducted by the Bureau of Labor Statistics. The current nonresponse follow-up approach uses a telephone interview which asks the potential respondent to provide an *estimate* of hours paid and hours at work. The new telephone-prompt approach requires interviewers to identify and trace an appropriate informant, and asks the informant to use the establishment records based hard data. The experiment (n = 1,200) was embedded in the HWS which collected data for the 1995 reference period. Results indicate that the prompting-script based method converts more nonresponding units than the current method *per hour*, and that the data collected by the prompting-script is more complete and accurate than those collected by the current method. These survey measures are further analyzed by looking into important attributes of informants (familiarity with the survey, and intention to comply), and their organizational characteristics (e.g., industry type, size, and multi-vs.-single-site). The final part of this paper examines the reasons for refusals which will increase our understanding of "hard-core holdouts" in establishment surveys

Chiu, Christina, Robertson, Kenneth, and **Tou, Albert** (2000), "An Evaluation of the Occupational Staffing Pattern and Wage Rates of State Government Workers," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The Occupational Employment Statistics (OES) survey, conducted by the U.S. Bureau of Labor Statistics, is an annual survey of 400,000 nonagricultural business establishments. The survey is designed to collect information about occupational employment and wages. Estimates of occupational employment and wages are produced using three years of data. In order to reduce the burden to respondents most establishments are sampled only once during a three year time period. State government

establishments, however, are included in the sample every year. Some of the survey's data collectors have conjectured that the occupational staffing patterns of state government workers are stable from year to year, and that the year to year changes in the occupational wage rates are explained by the rate of change in wages developed by the Bureau's Employment Cost Index survey. If both of these conjectures are true, then we do not need to collect these data each year. A variation of the chi-square statistic and the Bonferroni multiple comparison test will be used to evaluate the year to year change in occupational staffing patterns and wage rates, respectively, of state government workers. The evaluation will be based on 1996, 1997, and 1998 OES state government data.

Chun, Young and **Robertson, Kenneth** (1996), "Reducing Nonresponse in Business Surveys" *Proceedings of the Section on Survey Research,* American Statistical Association, 980-985.

A large scale randomized experiment (n = 6,000) conducted in 1995 by the Bureau of Labor Statistics showed that the combined use of advance letters and reminder/thank you letters has significantly increased the response rate in a typical <u>establishment</u> survey where the unit of analysis is a nonagricultural business. This finding confirms the results in the previous small experiment (n = 400) presented in the 1995 AAPOR conference, and fully complements the findings based on the household survey which have been accumulated by the Dillman school since the mid-1970s. We also learned that both early additional contacts helped us identify nonviable business units or refusals so that we can allocate resources up front in a more knowledgeable manner.

Clayton, Richard (1997), "Implementing Fully Electronic Data Collection In The Redesign Of The Current Employment Statistics Survey," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 295-297.

The Current Employment Statistics (CES) program is a monthly panel survey collecting employment and payroll data from 400,000 business establishments. Since the early 1980s, the CES has developed and implemented a range of automated data collection methods, including Computer Assisted Telephone Interviewing (CATI), Touchtone Data Entry (TDE), FAX, Electronic Data Interchange (EDI) and World Wide Wed (WWW) collection. These integrated systems and methods now handle monthly collection for over 240,000 single and small multi-unit establishments. Electronic Data Interchange (EDI) is used for the largest multi-unit firms. Beginning in July 1997, the CES is fielding a probability sample design. This network of automated methods, including systems,

procedures, staffing and cost structure, provides the basis for fully automated data collection under the new design. This paper describes the existing set ofautomated data collection methods and the changes necessary for the new design requirements, including identification of new births, deaths, solicitation, and ongoing collection and attrition prevention techniques needed to maintain high response rates over time in a fully automated, mixed mode collection environment.

Clements, Joseph, (2000), "Protecting Data in Two-Way Statistical Tables Using Network Flow Methodology," *Proceedings of the Section on Government Statistics,* American Statistical Association.

Cell suppression is a standard technique used by many federal statistical agencies to protect statistical tables. If a cell in a table reveals too much information about individual respondents' values, then it is suppressed when the table is published. To protect these cells (i.e., primary cells) from being resolved using other information in the table, additional cells, called complementary cells, must also be suppressed before the table is published. For even moderately sized tables, identifying an optimal complementary cell suppression pattern, that is, one that protects all of the primary cells and minimizes the amount of information lost, is a complex numerical problem. Because of this, heuristic techniques have been developed that attempt to minimize the cost function by protecting one primary cell at a time, iterating until all of the primary cells are protected. For two-way tables, heuristic techniques based on network flow formulations are preferred because they provide good results and they can be solved using numerically efficient algorithms such as the network simplex algorithm. In this paper, we examine some of the network flow formulations that have been proposed to solve the complementary cell suppression problem in two-way tables.

Cohen, Stephen (1997), "The National Compensation Survey: The New BLS Integrated Compensation Program," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 451-456.

The National Compensation Survey (NCS) is both a replacement for the Occupational Compensation Survey Program (OCSP) which estimates occupational pay levels by skill and also a program that integrates this new survey with two existing BLS compensation surveys: the Employment Cost Index (ECI) which measures quarterly change in employer compensation costs and the Employee Benefits Survey (EBS) which measures participation rates and gives details of employer provided benefits. A key design change from OCSP is to select a pps sample of occupations in each establishment matching skill level to generic definitions rather than

surveying a fixed list of benchmark occupations. The NCS uses a rotating panel design with three stages of selection in each panel. When fully integrated ECI will also be able to measure quarterly change in employer costs by occupational skill level in addition to industrial and occupational detail and EBS will link participation and provisions data with employer cost. This paper will focus on survey output differences and point out relevant statistical design issues that will be addressed in more detail in other papers at this meeting.

Cohen, Stephen, Wright, Tommy, Vangel, Mark, and Wacholder, Sholom, (2000), "Postdoctoral Research Programs in the Federal Statistical System," *Proceedings of the Section on Survey Research Methods,* American Statistical Association. .

Federal agencies use a variety of means to ensure exchange of expertise between the academic environment and the federal statistical system. For example, agencies encourage established researchers to apply for sabbaticals by funding efforts such as the ASA/NSF fellowship programs. The Postdoctoral Research Programs in the federal agencies are designed to attract recent PhD graduates. In either program, researchers gain access to data that is not easily accessible outside the agencies and have the opportunity to do research which enhances both their career development and the federal sector. Program participants work on a variety of applied scientific/policy/data collection problems. The goal is to attract talented people to consider the federal sector for a career and to build ties between the federal statistical community and researchers who ultimately plan careers in the academic sector. Panel participants will discuss the potential benefits of developing these programs, benefits of the program to the participants, different ways of administering postdoctoral research programs, and implications of their effect on the federal statistical system.

Shettle, Carolyn, Ahmed, Susan, **Cohen, Steve**, Miller, Renee and Waite, Preston (1997), "Improving Statistical Reports From Government Agencies Through The Reports Review Process," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Statistical agencies use a variety of means to assure that their reports meet high quality standards in terms of relevancy, accuracy, and objectivity. One key component in most statistical agencies, is the reports review process. Procedures for conducting these reviews vary considerably among agencies. Panel participants will describe the procedures used in their agencies (Bureau of Labor Statistics, Census, Energy Information Administration, National Center for Education Statistics, Science Resources Studies/National Science Foundation) and will discuss the

implications of these procedures for report quality, staff morale, and timeliness.

Rips, Lance, **Conrad, Frederick** and **Fricker, Scott,** (2000), "Unraveling the Seam Effect," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Panel surveys sometimes ask respondents for data from several different intervals within a longer reference period. For example, the Survey of Income and Program Participation interviews its respondents every four months, but it asks about each month separately within the four-month interval. Findings from such surveys often show larger changes from one month to the next when the data come from two different interviews than from the same interview. We have studied this seam effect experimentally in a setting that allows us to control the information that respondents should report. The results of the experiments are consistent with a theory in which the seam difference is due to two factors: (a) respondents' forgetting information within the response interval, and (b) their bias in reporting when they can no longer remember correct answers.

Conrad, Frederick, Blair, Johnny and Tracy, Elena (1999). "Verbal Reports are Data! A Theoretical Approach to Cognitive Interviews." *Proceedings of the Federal Committee on Statistical Methodology Research Conference.*

The use of verbal reports to pretest questionnaires (cognitive interviews) is the most tangible outcome to date of the dialogue between cognitive psychology and survey methodology. Cognitive interviews are a standard survey pretesting tool yet they rarely exploit the theory and body of knowledge about verbal report methods. For example it is believed that if people are not aware of a thought process, they cannot verbalize it. In addition, there is evidence that verbalizing certain processes can affect the process being reported. When these are overlooked, it threatens the validity of verbal reports. We have developed a two part technique for collecting and analyzing verbal reports in cognitive interviews that takes this into account. One part concerns collecting verbal reports – the administration of cognitive interviews. The second part concerns the analysis and interpretation the verbal reports. In the data collection part, cognitive interviewers use certain generic probes when a verbalization indicates the respondent is aware, of but has not reported, useful information. In the interpretation and analysis part, coders assign segments of the verbal reports to a problem taxonomy in which a set of problem classes can be occur throughout the stages of the response process. We advocate using

an approach like this to increase the validity and objectivity of cognitive interview data. Preliminary data suggests that it is promising.

Conrad, Frederick, Brown, Norman and **Dashen, Monica (1999). "**Estimating the Frequency of Events from Unnatural Categories." *Proceedings of the Section on Survey Research Methods,* American Statistical Association. .

"How often do you do light or moderate activities for at least 10 minutes that cause only light sweating or a slight to moderate increase in breathing or heart rate?" This is a hard question because it asks about an event category ("light or moderate activities …") which seems to be at odds with the way most respondents think about events. We argue that the danger in asking respondents about the frequency of such categories – we call them unnatural categories – is that relevant episodes may not come to mind, leading to underreporting. We explored this in two studies. In both, participants studied a list of words and were then asked to estimate the number that were members of either unnatural or natural categories. In the first experiment, the unnatural category group was asked how many of the words they had studied contained particular *properties* (e.g. shiny, smelly, round); the group tested on more natural categories was asked how many words were members of *taxonomic categories* (e.g. furniture, mammals, fruit). Both groups underestimated actual frequency but it was far more extreme for those answering about properties. In the second experiment, both groups were tested on properties but one of these groups, the *instance + property* group, studied the properties along with the individual words (e.g. milk-white) to see if by promoting encoding of the properties we could render them more natural. Response-time patterns indicated that this was the case: *instance-only* estimates took three times as long as *instance + property* estimates and involved far more underestimation of actual frequency. We recommend decomposing unnatural categories into their natural parts and asking separate questions about each.

Conrad, Frederick G. and **Schober, Michael F.** (1999) "Conversational Interviewing and Data Quality." *Proceedings of the Federal Committee on Statistical Methodology Research Conference. .*

Standardized interviewing is widely practiced because it promises to reduce interviewer related error and because it is cheap. Yet the technique cannot guarantee uniform understanding of questions and, thus, may reduce data comparability. Conversational interviewing may standardize the meaning of questions by allowing interviewers to clarify survey concepts, but it

cannot guarantee uniform interviewer behavior. We discuss four experiments (three in the laboratory where respondents answer on the basis of fictional scenarios allowing us to directly assess accuracy) and one in the field (in which respondents are interviewed twice so that response change and explanation allow us to indirectly assess accuracy). We have found that conversational interviewing improves response accuracy when respondents' circumstances are atypical (e.g. Should a lamp purchase be counted as a furniture purchase?) but requires additional time to clarify concepts. The more ways in which clarification can be provided the more accurate are responses and the longer interviews last. Respondents are not always willing to seek clarification when they need it, though we were able to overcome this in one experiment by reducing the social and cognitive demands of asking for clarification. Instead of interacting with an interviewer respondents interacted with a computer and clicked a mouse on highlighted text to obtain clarification; when respondents were told that clarification was essential for accurate responding they frequently obtained it. We conclude that researchers have the option of keeping interviews short and risking some misunderstanding or investing more time to be sure questions are understood as intended.

Schober, Michael F., **Conrad, Frederick G.,** and **Fricker, Scott S.** (2000). **"**When and How Should Survey Interviewers Clarify Question Meaning?" *Proceedings of the Section on Survey Research Methods,.* American Statistical Association.

 Conversational interviewing, in which interviewers and respondents work together to make sure questions are understood as intended, can help respondents answer more accurately than strictly standardized interviewing. It takes longer and can be costly. Here we measured response accuracy and interview length for three kinds of partially conversational interviewing, which resemble current practice in some organizations. Census Bureau interviewers telephoned paid laboratory respondents, who answered factual questions from ongoing government surveys on the basis of fictional scenarios. Interviewers either (1) read scripted definitions of question concepts when respondents explicitly asked for clarification; (2) used their own words to present official definitions of question concepts when respondents explicitly asked for clarification; (3) presented scripted definitions whenever they deemed it necessary, even if respondents hadn't explicitly requested clarification. For all three partially conversational techniques responses were reliably more accurate than for strictly standardized interviews, and interviews took reliably longer; compared to more fully conversational interviews (in which interviewers could present definitions in their own words whenever they thought this might help) responses were less accurate and interviews shorter. Results suggest that response accuracy improves whenever respondents get help,

whether or not they ask for it explicitly; but if it is left only up to respondents to ask for help, they often won't.

Conrad, Frederick and Schober, Michael (1998), **"A Conversational Approach to Self-Administered Questionnaires,"** *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 962-967.

Conversational survey interviews can produce more accurate responses to factual questions than standardized interviews under some circumstances (Schober and Conrad, 1997; Conrad and Schober, under review). The current study examines how accuracy is affected by (1) the way respondents can obtain clarification, and (2) whether or not they are instructed to obtain it. We simulated a standardized interview and four kinds of conversational interviews in a CASI environment. Questions were presented on a computer and respondents answered based on fictional scenarios, enabling us to determine the correct answers. One group of respondents could not obtain any clarification and four others could: they either clicked their mouse on highlighted text for a definition or, if they were inactive, the computer offered to present one. Half were told that in order to respond correctly they would need to obtain definitions; the others were only told that definitions were available if they wanted to use them. For those cases in which a definition was potentially helpful, respondents were only accurate when help was available and they were instructed to use it. The results suggest that CASI interactions that embody certain features of conversational interviewing can improve response accuracy and that respondents may be more willing to ask for clarification by clicking a mouse click than by asking an interviewer.

Schober, Michael and **Conrad, Frederick** (1998), "Response Accuracy When Interviewers Stray From Standardization" Proceedings of the Section on Survey Research Methods, American Statistical Association, 940-945.

Strictly standardized interviewing can lead to poorer response accuracy than more conversational interviewing methods (Schober and Conrad, 1997; Conrad and Schober, under review). In these studies, we trained one group of interviewers to use Fowler and Mangione's (1990) version of standardized interviewing, and another group to use one version of conversational interviewing, reading questions exactly as worded and then working together with respondents to make sure they interpreted each question as intended. In the current study we examine response accuracy under the interviewing techniques that are actually practiced by 11 interviewers from the telephone center that conducted the Schober and Conrad (1997) study. The 21 respondents answered on the basis of

fictional scenarios, so that we could measure response accuracy directly. Half of the scenarios ("straightforward") corresponded to particular survey questions in an obvious way; the other half ("complicated") described situations that did not map onto the survey questions straightforwardly. Interviewers deviated from strictly standardized procedures on 19.8% of the questions. These interviewer deviations didn't affect response accuracy for straightforward mappings, which was nearly perfect. But for complicated mappings, interviewer deviations improved accuracy to 79.4%, compared to a 22.8% accuracy rate when interviewers followed strictly standardized procedures.

Conrad, Frederick and Blair, Johnny (1996), "From Impressions To Data: Increasing The Objectivity Of Cognitive Interviews," *Proceedings of the Section on Survey Research Methods,* American Statistical Association, 1-9.

The use of "think aloud" protocols is the most tangible application of techniques from cognitive psychology to survey methodology -- the so-called "cognitive interview." Think aloud methods can be used systematically but in questionnaire pretesting, the technique is usually used differently by different practitioners and relies heavily on the intuitions of the particular practitioner. This paper explores an approach to the cognitive interview that promises to increase its consistency and to provide a theoretical basis for the method. The central components of our technique are (1) a procedure for eliciting the intent of each question from question authors, and (2) a taxonomy of common respondent problems used to code think aloud protocols. By comparing the intent of the question and the content of the protocol, the researcher has relatively objective criteria for identifying where a respondent's interpretation and response behavior differ from the question author's intent, thus identifying question problems. We present a study that tests the reliability of the approach and compares it to the conventional use of the technique.

Conrad, Frederick and Schober, Michael (1996) **"**How Interviewers' Conversational Flexibility Affects the Accuracy of Survey Data,**"** *Proceedings of the Section on Survey Methods Research,* American Statistical Association, 883-888.

Standardized survey interviewing is widely advocated in order to reduce interviewer-related error (e.g. Fowler & Mangione, 1990). But Suchman and Jordan (1990, 1991) argue that standardized wording may decrease response accuracy because it prevents the conversational flexibility that respondents need in order to understand what the survey designers have intended. We evaluated these competing positions – *standardized* versus *flexible* interviewing approaches -- in a laboratory experiment.

Professional interviewers asked the respondents questions from three large, government surveys, using either standardized or flexible interviewing techniques. The respondents provided answers on the basis of fictional descriptions enabling the experimenters to measure response accuracy. Accuracy and certain characteristics of the interaction were compared under standardized and flexible interviewing conditions. The two interviewing techniques led to virtually perfect accuracy when the concepts in the questions clearly mapped onto the fictional situations. When the mapping was less clear, flexible interviewing increased accuracy substantially, from 27% to 87%. We discuss theoretical and practical implications and propose which circumstances justify the use of either interviewing technique.

Conrad, Frederick G. (1996), "Using Expert Systems To Model And Improve Survey Classification Processes" In Lyberg, L., Biemer, P., Collins, M. DeLeeuw, E., Dippo, C. Schwarz, N. & Trewin, D. (Eds.), *Survey Measurement and Process Quality*. New York, John Wiley & Sons.

In order to analyze most survey responses, the data must be assigned to categories. When the response task does not include classification (as in answering open ended questions) responses are classified by coding specialists, usually after the raw data are collected. This chapter distinguishes between simple and complex classification and argues that expert system software is particularly good at supporting the second of these. Simple classification usually involves locating the response in some sort of dictionary where it is explicitly mapped to a particular category. The literature about automated coding of this sort is first reviewed. Then complex classification is presented as a task requiring the coding specialist's expertise about the content area. This is illustrated with two prototype expert systems that embody specialists' knowledge of occupational classification in the Occupational Compensation Survey Program and the review of commodity substitutions in the Consumer Price Index. Finally, some of the practical considerations of implementing expert systems in survey organizations are presented.

Katz, I., **Conrad, F.** and **Stinson, L.** (1996), "Questionnaire Designers versus Instrument Authors: An Investigation of the Development of CASIC Instruments at the Bureaus of Labor Statistics and of the Census," Paper presented at the Inter-CASIC Conference, San Antonio, Texas.

Much of the literature on CASIC development focuses on available technologies and their promise or on tools and techniques available to instrument authors (i.e., programmers). Less is reported about the

interaction between programmers and the staff (questionnaire designers) who develop and specify questionnaire that the CASIC instrument is to embody. We investigate the process of designing and developing CASIC through (a) interviews with authoring staff and questionnaire design staff at BLS and Census and (b) a questionnaire, distributed to these staff, concerning the CASIC development process.

Crankshaw, Mark and **Stamas, George,** (2000), "Sample Design in the Job Opening and Labor Turnover Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The Job Opening and Labor Turnover survey (JOLTS) is a new program being fielded by the Bureau of Labor Statistics to compile measures of job openings and employee turnover for nonfarm businesses using a sample of 16,000 establishments. The job opening, employee turnover (hires and separations), and employment data are collected on a monthly basis. Sample members will rotate out of the sample after participating for 18 months. Estimates will be produced for broad industry groups and Census region. In this paper we will discuss JOLTS sample design, sample coordination with other BLS surveys, JOLTS estimators, and issues arising from sample selection using the Standard Industrial Classification (SIC) system as a strata identifier and estimating for North American Industrial Classification System (NAICS) sector. The JOLTS sample was coordinated with other BLS establishment surveys using permanent random numbers to minimize overlaps between surveys. Sample was selected by industry divisions defined on SIC because the conversion from SIC to NAICS was not yet completed, however, the NAICS conversion will be complete before we begin publishing an experimental series.

Dashen, Monica and **Fricker, Scott** (1998), "Taking Different Perspectives on a Survey Question," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Omissions (the failure to report all relevant items) are a potential source of measurement error. While omissions are often attributed to forgetting, they may also be due to the respondent's tendency to approach questions with purposes or perspectives, which differ, from the designer's intent. Drawing from a body of research on perspectives, we attempt to understand how omissions occurred in some questions from the Telephone Point of Purchase survey. These questions pertain to beverage and day care expenditures. We asked respondents to read a list of words from one to two perspectives and recall them later (regardless of the perspective taken during the study phase). For example, some respondents were asked to

study a list of Tea & Coffee items list in terms of expenses needed to make coffee, while others studied it in terms of items needed to accompany coffee. Our preliminary findings suggest that a word's significance in terms of a given perspective determined the probability that it would be recalled. For example, respondents in the coffee making group tended to recall more items related to coffee making (e.g. water and coffee) compared to those in the accompany group. These results suggest perspectives can affect the respondents' interpretation of the question and the retrieval of information.

Dashen, Monica, (2000), "Improving Purchase Recollection," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Retrospective reports about personal events are often collected by a survey long after the event in question. It is well known that people forget personal events over time, which can affect the data quality. Yet little is known about how to improve recollection. Accordingly, two studies explored how well survey respondents were able to remember frequently-purchased items and examined ways to improve purchase recollection. In both studies, participants recorded their own purchases for two weeks and then returned, either immediately, one week, two weeks or four weeks later for a recall test of their own purchases. Study 1 was designed to measure the effects of time and salience on memory for purchases. The results indicate that memory deteriorated over time, as indicated by low match rates and high intrusion (or false report) rates. In Study 2, some participants were given memory enhancers to improve their purchase recollection. The match rate data suggest that participants whose memories had been enhanced reported more purchases correctly than did those participants whose memories had not been enhanced. The results of both studies underscore the importance of understanding memory-based response error in surveys, as well as devising and implementing techniques to improve memory.

Dashen, Monica and **Sangster, Roberta L.** (1997), "Does Item Similarity And Word Order Influence Comparative Judgments?," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Respondents in surveys are often asked to make comparative judgments between two items, i.e., A and B. Evidence suggests that comparative judgments are influenced by the order of presentation (A to B vs. B to A). It has been posited by Wanke, Schwarz and Noelle-Newman (1995) that when asymmetric judgments occur they are due to attention to unique features of A and B. This theory implies that when item similarity and word order is manipulated, the likelihood of an asymmetric judgment increases as

the number of unique features between two items increases. This paper seeks to test this implication and replicate the Wanke et al (1995, study1) finding. Three-hundred respondents completed one out of six ballots of a self-administered survey. Respondents were asked two questions pertaining to (1) empathy for college related problems from male and female teachers or friends and family and; (2) job market preparation by high and low contrast majors. The direction of the comparison was manipulated by the wording of questions. The present work failed to replicate Wanke et al. (1995) and provide evidence for asymmetric judgments as a function of similarity. Theoretical and applied issues will be discussed.

de Wolf, Virginia A. (1997), "The "Interagency Confidentiality and Data Access Group," *Proceedings of the Section on Government Statistics*, American Statistical Association.

The Federal Committee on Statistical Methodology (FCSM) was organized by the Office of Management and Budget in 1975 to investigate issues of data quality affecting federal statistics. The work of FCSM is conducted through subcommittees that are organized to study particular issues. Typically, the results of the deliberations of these subcommittees are contained in a FCMS series called Statistical Policy Working Papers (SPWPs). In 1994 FCSM released SPWP 22, 'Report on Statistical Disclosure Limitation Methodology.' SPWP 22 contained a primer of basic methods, a summary of agency practices, a set of recommendations for disclosure limitation practices, and a research agenda. In 1995 the Interagency Confidentiality and Data Access Group (ICDAG) was formed to promote the goals and objectives of SPWP 22 and to foster increased cooperation and sharing on statistical disclosure methods. In 1997, FCSM wanted to develop a mechanism to facilitate communication and cooperation among agencies and ICDAG became its first "Interest Group." This paper describes the objectives, membership, as well as the goals of ICDAG.

Dippo, Cathryn S., Gillman, Daniel W. (1999), The Role of Metadata in Statistics," *Paper presented at UN/ECE Work Session on Statistical Metadata* .

Metadata The term is used by many people speaking about many different things. Just what does the term mean with respect to official statistics? While the dictionary definition "data about data" is concise and accurate, it lacks the specifics and context needed to communicate meaning. So, a few years ago, we developed the following definition:"Statistical Metadata is descriptive information or documentation about statistical data, i.e. microdata,macrodata, or other metadata. Statistical Metadata facilitates

sharing, querying, and understanding of statistical data over the lifetime of the data." This definition is also fairly concise and accurate; moreover, it provides some context. But is it sufficient to convey meaning to a diverse set of users such that their comprehension of the term is equivalent? Probably not.. Thus, our goal in this paper is to indicate the breadth of meaning associated with the term metadata in the context of official statistics and the agencies that produce them.

Gregg,Valerie and **Dippo, Cathryn S.** (1999), *"FedStats: Partnering to Create the National Statistical Information Infrastructure of the 21st Century," Proceedings of the Section on Government Statistics,* American Statistical Association.

Two years ago, the Office of Management and Budget's Interagency Council for Statistical Policy opened the FedStats WWW site (http://www.fedstats.gov)--a "One-Stop Shop for Federal Statistics". The FedStats Interagency Task Force continues to design and develop new approaches to make it easy for users to find Federal statistical information, without having to know which of the 70+ agencies holds the data and information users seek. The Task Force keeps FedStats a dynamic, award winning website by working together to develop new features. FedStats lessons learned contribute to defining longer-term computer science and information technologies research requirements for realizing the Federal statistical community's vision for a National Statistical Information Infrastructure. To achieve this vision, FedStats agencies are undertaking research partnerships by collaborating with academic computer scientists, social scientists, statisticians; and, our local and state partners. Research partnerships seek to bring the latest technologies to bear on data integration, knowledge discovery, data validation and visualization, data retrieval, storage and archiving, computational intensive statistical modeling, on-line collaborative problem solving environments, and universal access.

Dippo, Cathryn S. and Hoy, Easley (1997), "Providing Metadata to Survey Staff Via Internet," Presented at ISI 51st Session – Istanbul,Turkey.

Sharing information about the survey process – from conceptualization through planning, design, development, and implementation to validation— is critical to a successful survey operation. Universal access to the Internet by all workers involved in a project now makes it possible to change the process for creating and distributing process information.

Dippo, Cathryn S. and Tupek, Alan (1997), "Quantitative Literacy: New Website for Federal Statistics Provides Research Opportunities," *D-Lib Magazine*, December 1997

> The public release of FedStats in May 1997 has been a phenomenal success by providing one-stop shopping for all federal statistics. Continued interdisciplinary research may not only permit users to find the statistics faster and easier but may also help users understand how best to use the statistics. FedStats may, therefore, become an icon for improving the quantitative literacy of all Americans.

Dorfman, Alan H. (1999), "Issues in the Analysis of Complex Surveys"," *Proceedings Book2 Topic 67 (Bulletin of the International Statistical Institute)*.

> Data from a survey of a population can be used for two goals: (1) description *of* the population (i.e. estimation of such numerical characteristics as the total values of a given variable); (2) inference *from* the population to the process which generated it. Deming (1950, Chap 7) refers, on the one hand, to descriptive (or *enumerative*) studies, and, on the other, to *analytic* studies, which are aimed at "the causes of patterns and variations…". According to Skinner, Holt, &Smith 1989 [henceforth SHS], p.1, "… the analytic use takes us beyond the summary measures embodied in description and into the causal explanation of the processes that underlie the descriptive measures." Thus, analysis is concerned with uncovering the causal relations between things, and the relations between corresponding variables, not only in the population sampled, but in like populations. A derivative, but not entirely coincident, definition of the analysis of surveys, understands the population to be characterized by a statistical model, and the goal to be the estimation of the parameters of the model (SHS, p.5). This paper explores several questions that arise from the tension between these two notions of survey analysis.

Dixon, John, (2000), "The Relationship Between Household Moving, Nonresponse, and the Unemployment Rate in the Current Population Survey," *Proceedings of the Section on Government Statistics,* American Statistical Association.

> In the Current Population Survey, a household survey from which labor force estimates are produced, selected housing units remain in sample during a 16-month period. The households are interviewed during the first 4 and last 4 months of this period. During this time, the household occupying a sample housing unit may change. Matching households between months allows an analysis of the relationship between whether a household moves and estimates of the employment rate. Many households

move during the 16 months they are in sample. Since change in employment may be related to the household's decision to move, the estimates of employment status may be affected. "Inmovers" don't completely make up for the number of "outmovers" so their relative effect may not be offset. The differences in response rates can also affect estimates. The current study examines the nature of this relationship through an analysis of the characteristics of movers and the resulting effect on labor force estimates.

Dorfman, Alan H. (1999), ″The Stochastic Approach to Price Indices," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

More than most statistical concepts, price indices seem arbitrary and elusive. Three rather different approaches have been developed for choosing among alternative price index formulae: (1) the test approach, (2) the economic approach, and (3) the "stochastic" approach through statistical models. The stochastic approach was investigated and championed in the last century and beginning of this, most notably by Jevons and Edgeworth, but after criticism by Keynes, and with the rise of the economic approach, the stochastic approach fell by the wayside. It is perhaps not an accident that no article on price indices has appeared in JASA for many decades. To calculate price indexes, data on "the same item" (in most cases, actually a collection of items narrowly defined) is collected across time periods. Is it feasible to use statistical modeling of such "quasi-longitudinal" data for characterizing price indexes? A state space model of price data is suggested, yielding a consumer price index defined in terms of the parameters of the model. Feasibility of the new index is shown using scanner data for canned tuna. The stochastic approach, brought up to date, brings new insight to the question of price indices.

Dorfman, Alan H., (2000) "Non- Parametric Regression for Estimating Totals in Finite Populations," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Nonparametric regression provides computationally intensive estimation of unknown finite population quantities. Such estimation can be more efficient than inference tied to design-probabilities (in design-based inference) or to parametric regression models (in model-based inference), and, furthermore, gives deeper insight into traditional methods such as poststratification.. A nonparametric regression based estimator for finite population totals is introduced. Its utility is supported by determination of its asymptotic bias and variance and by simulation studies on real and artificial data.

Dorfman,Alan, Leaver,Sylvia; and **Lent, Janice (1999).** "Some Observations on Price Index Estimators," Statistical Policy Working Paper 29 - Part 2 of 5, pages 56-65.

Price indexes can be divided into two broad classes, the *superlative* indexes, and the non-superlative indexes. Superlative indexes in theory approximations a true cost-of-living index (which it is impossible to directly calculate), and tend in practice to be very close to each other, so that anyone of them can be taken as representing the class. Non-superlative indexes (which are the ones actually used in practice, because of the timeliness with which their sample based estimates can be produced) deviate in theory from the cost of living index, and in practice from the superlative indexes. Particular forms of-non-superlative indexes tend to lie above the superlative indexes, others below. The (positive or negative) gap between a non-superlative index and the cost of living index [has been characterized as] is its *substitution bias* or *substitution effect.*

Implicit is the notion that the indexes are *population indexes*, representing the totality of transactions of a given sort in a given economy. It seems natural to estimate the magnitude of the substitution effect of a given index by retrospectively measuring the distance between it and a superlative index. However, such measurements are necessarily made on *sample estimates* of the corresponding indexes. We show that the relationships among sample-based indexes, and between them and population indexes, are not necessarily a straightforward reflection of the relation between population indexes. Thus estimates of the direction and magnitude of the substitution effect are more imprecise than has previously been supposed.

Wang, Suojin, **Dorfman, Alan H.,** and Chambers, Raymond (1999)"*M*aximum Likelihood Under Informative Sampling*," Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Informative sampling occurs when the probability of inclusion in sample depends on the value of the survey variable. Most sample survey inference methods depend on the method of sampling being uninformative. A general methodology for likelihood based survey inference, based on application of the Missing Information Principle (MIP; Orchard and Woodbury, 1972), which allows for informative sampling is set out in Breckling, et al (1994). Unfortunately, this approach does not specifically model the distribution of the sample data, and so it is not obvious that maximum likelihood based on it leads to exactly the same inference as a more direct approach which builds upon this sampling distribution. Consequently it is of interest to demonstrate the equivalence of the ``MIP-based" and more direct approaches to maximum likelihood for sample data obtained via an informative sampling scheme. This paper does this in the context of a

particular method of sampling, called array sampling. Extensions to the theory are also explored.

Dorfman, Alan H. and **Valliant, Richard** (1997), "The Hajek Estimator Revisited," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 760-765.

The Hajek estimator of a population mean was introduced by Hayek (1971) as a competitor to the standard Horvitz-Thompson unbiased estimator, in the context of a paradoxical challenge to pps sampling by D. Basu (1971). It is approximately design unbiased and is widely highly regarded; cf. Sarndal, Swenson, Wretman (1992, pp.182-183). A recent Theorem of Royall (1992) which generalizes and extends earlier work on balanced sampling has a number of implications for survey thought and practice. Under certain forms of weighted balance that are obtained in expectation under pps sampling, the (model-based) BLU estimator of the mean reduces to the Hayek estimator. On the other hand, under (even minor) deviations from balance, which are quite likely to be met under pps sampling, the Hayek estimator exhibits a severe bias, which renders it unfit for use, especially when a reasonable working model and auxiliary data are available.

Short, Kathleen, **Doyle, Patricia,** Hernandez, Donald, Naifeh, Mary, Johnson, David, **Garner, Thesia** and Rozaklis, Patricia (1999), " Measuring Poverty: Questions, Approaches,and Findings," *Proceedings of the Section on Social Statistics,* American Statistical Association.

Session description: This presentation will summarize new findings described in the first report presenting alternative poverty measures. Released in the spring of 1999, that report will present an array of alternative poverty measures based primarily on the recommendations of the National Academy of Sciences. As a result of the release of that report, and subsequent discussion and comment, additional research will be underway. Led by an interagency committee formed for this purpose, the Census Bureau and the Bureau of Labor Statistics will continue to refine the proposed measures. Authors will discuss relevant issues, report on work undertaken to address those issues, and outline plans for additional research and data collection for the longer term.

Echols, Carrae, Robertson, Kenneth, and **Tou, Albert** (1999), *"*Developing Estimators for Use with Multiple Years of Grouped Data in the Occupational Employment

Statistics Survey," *Proceedings of the Section on Government Statistics,* American Statistical Association.

The Occupational Employment Statistics (OES) survey is an employment and wage survey of nonfarm business establishments conducted by the Bureau of Labor Statistics. Occupational employment and wage rate data are collected in a continuous 3-year cycle. The employment of each occupation is reported by coding employees into one of 11 contiguous, non-overlapping wage intervals. Note that the exact wage of individual workers is not reported. In this paper we evaluate alternative estimators used to calculate the mean and median wage rates of each occupation based on multi-year data sets. Exact wage data collected by the Bureau's Compensation and Working Conditions (CWC) surveys are used to evaluate these estimators by grouping the CWC data into wage intervals. Afterwards, alternative mean and median wage rate estimators are used to estimate wage rates for each occupation based on the grouped data. "True" mean and median wage rate values are calculated for each occupation by using the exact wage data reported by the CWC surveys. The wage rate estimates produced by each estimator are then compared to these corresponding "true" wage values.

Eltinge, John, (2000), "Implications of Model Validation Criteria for the Performance of Small Domain Estimation Methods," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

In work with sample survey data, small domains are subpopulations from which one has collected relatively few observations. Due to sample size limitations, customary design-based estimation and inference methods may perform relatively poorly when they are applied directly to data from these small domains. Consequently, the sample survey literature frequently uses model-based approaches to develop small domain estimation methods, and to evaluate the operating characteristics of these methods. This paper examines the extent to which some customary model validation criteria can be linked with the operating characteristics of the resulting small domain estimators. Special attention is devoted to cases involving moderate lack of fit.

Park, Inho, and **Eltinge, John,** (2000), "The Effect of Cluster Sampling on the Covariance and Correlation Matrices of Sample Distribution Functions," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

This paper uses a superpopulation to examine the effect of cluster sampling on the covariance and correlation matrix of sample distribution functions. A simplified two-stage sampling design leads to a covariance matrix approximation which is an extension of the well-known formula (e.g., Cochran, p. 242) for the univariate sample mean. The resulting estimation method is then be applied to a multistage stratified sampling design in an ad hoc way to provide a model-based estimator of the covariance matrix of sample distribution functions and the approximate distribution of sample quantiles. The proposed methods are applied to medical examination data from NHANES III.

Yansaneh, Ibriham, and **Eltinge, John,** (2000), "Design Effect and Cost Issues for Surveys in Developing Countries," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Operating characteristics of surveys, such as design effects, cost, and response rates, are frequently used for planning purposes by international agencies that support the design and implementation of surveys in developing and transition countries. In this paper we describe a framework for comparing operating characteristics of surveys across different types of surveys and different countries or regions. We will extend the work of Verma and Le (1996) on design effects by considering other surveys and the decomposition of design effects into more detailed components associated with stratification, clustering, and unequal-probability sampling. The decomposition of the other operating characteristics into relevant components will also be discussed.

Eltinge, John (1999), "Evaluation and Reduction of Cluster-Level Identification Risk for Public-Use Survey Microdata Files," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Design-based analyses of cluster sample data generally require the use of primary sample unit (PSU) labels, or of closely related replicate weights. However, for cases in which PSUs are identical to counties or groups of counties, inclusion of PSU labels in public-use data can increase the risk that some PSUs will be identified. This is problematic because statistical agency policies frequently prohibit the public release of information that would allow the identification of small or medium sized PSUs. In addition, identification of a given primary unit can increase the risk of identification of constituent sample elements, e.g., persons or establishments. This paper discusses methods for the evaluation and reduction of PSU-level

identification risk. Principal emphasis is placed on methods that modify the original PSU label structure, but still permit the computation of approximately design unbiased variance estimators. The proposed methods are applied to data from the U.S. National Health Interview Survey (NHIS).

Heo, Sunyeong and **Eltinge, John** (1999), "The Analysis of Categorical Data from a Complex Sample Survey: Chi-squared Tests for Homogeneity Subject to Misclassification Error*," Proceedings of the Section on Survey Research Methods,* American Statistical Association.

In the analysis of categorical data, if misclassification errors exist, then estimated cell probabilities may biased and standard Pearson chi-squared tests may have inflated true type I error rates. This paper considers methods to evaluate the power of chi-squared tests for homogeneity with complex survey data subject to misclassification errors. Three cases are considered: adjustment with a known misclassification matrix; adjustment with an estimated misclassification matrix; and no misclassification adjustment. The proposed methods are applied to the data from the Dual Frame National Health Interview Survey (NHIS)/Random-Digit-Dialing (RDD) Methodology and Field Test Project.

Lee, Sangrae and **Eltinge, John** (1999), *"*Diagnostics for the Stability of an Estimated Misspecification Effect Matrix *," Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Analyses of sample survey data frequently use quadratic-form test statistics. Applications include comparisions of subpopulation means or proportions, simultaneous confidence bounds for a parameter vector, or goodness-of-fit chi-squared tests. One approach is to base the quadratic-form statistics on the inverse of an associated design-based covariance matrix estimator. This Wald-type approach works well for many applications, but in some cases, it can involve an unstable covariance matrix estimator and thus can inflate type I error rates. An alternative is to use a covariance matrix estimator closely related to first-order Rao-Scott adjustments. This paper develops diagnostics to compare the relative stability of the Wald and first-order Rao-Scott statistics. The diagnostics are based on the eigenvalues of an estimated misspecification effect matrix. Simulation methods are used to develop reference distributions and critical values for these diagnostics. The proposed methods are applied to data

from the U.S. Third National Health and Nutrition Examination Survey (NHANES III).

Park, Inho and **Eltinge, John** (1999), "Fitting Complex Survey Data to the Tail of a Parametric Distribution," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

For simple random samples, one sometimes estimates the mean and standard deviation of a normal population from regression of observed quantiles on the corresponding standard normal quantiles. This paper considers an extension of this idea to data collected through a stratified multistage sample survey. Principal attention is devoted to fitting a parametric model to the tail of an underlying superpopulation distribution. Direct application of ordinary least squares and generalized least squares methods lead to point estimators of the superpopulation parameters; associated variance estimators; and related goodness-of-fit test statistics. Separate consideration is given to methods based on constrained misspecification effect matrices. The proposed methods are applied to medical examination data from the U.S. Third National Health and Nutrition Examination Survey (NHANES III).

Parsons,Van and **Eltinge, John** (1999) "Stratum Partition, Collagpse and Mixing in Construction of Balanced Repeated Replication Variance Estimators," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The National Health Interview Survey (NHIS) uses a stratified multistage sample design, with stratification carried out separately within each state. For some analyses, it is desirable to approximate the true sample design with a simple two-primary-units-per-stratum design. For example, this simplified design allows the use of standard replicate half-sample variance estimators. This paper discusses the use of three methods to construct this two-per-stratum approximation within self-representing strata. First, stratum partition assigns most or all of the sample units in a stratum to a pair of pseudo-primary sample units. Second, stratum collapse is used to group together selected singleton primary units, as well as residual units from the abovementioned stratum partition. Third, stratum mixing allows one to reduce the effective number of strata without exacerbating customary bias problems induced by stratum collapse. Some state-level analysis work limits the number of replicates, and thus the number of effective strata, due to the computational burden associated with the use of

a large number of replicates. This paper closes with a discussion of trade-offs encountered in use of the abovementioned three methods.

Ernst, Lawrence and **Paben, Steven,** (2000), "Maximizing and Minimizing Overlap When Selecting Any Number of Units per Stratum Simultaneously for Two Designs with Different Stratifications," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

A number of procedures have been developed, beginning with the work of Keyfitz, for maximizing or minimizing the overlap of sampling units for two stratified designs. Certain overlap procedures have been developed for use when the two samples may be selected simultaneously. They generally produce a better overlap than procedures developed for sequential selection applications or are computationally more efficient. One of these procedures (Ernst 1996) is applicable when the stratifications for the two designs may be different, but is restricted to one unit per stratum designs. A second procedure (Ernst 1998) has no restrictions on the number of sample units per stratum, but requires that the designs have identical stratifications. In the present paper we present a simultaneous overlap procedure without the restrictions in the two previous papers; that is, there are no restrictions on the number of sample units per stratum, nor is there a requirement that the two designs have identical stratifications. This procedure, like the two previous procedures, requires the solution of a sequence of transportation problems and produces an optimal overlap. Potential applications are discussed.

Ernst, Lawrence R. (1999), The Maximization and Minimization of Sample Overlap Problems: A Half Century of Results," *Bulletin of the International Statistical Institute*, Proceedings Tome LVII, Book 2, 293-296.

Many procedures have been developed in the last half century, beginning with Keyfitz's (1951) pioneering work, to maximize or minimize the expected number of units retained in sample when a new sample is selected with selection probabilities that are different than those used to select the initial sample.. In this paper we discuss the properties of more than a dozen overlap procedures. For example, certain procedures are usable only for one sample unit per stratum designs, while other procedures can be used for designs for which there are a large number of sample units per stratum. Some procedures require identical stratifications for the designs being overlapped, while others do not. Some procedures do not work properly if used in two successive redesigns. Certain procedures use linear programming to produce a better overlap at the cost of additional

computational complexity. Some recently developed overlap procedures, with mostly desirable properties, can be used only when the samples for the designs being overlapped are selected simultaneously.

Ernst, Lawrence R., Valliant, Richard and **Casady, Robert J.** (1998), **"**Permanent and Collocated Random Number Sampling and the Coverage of Births and Deaths," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 457-462.

Permanent random number (PRN) and collocated random number (CRN) sampling are practical methods of controlling overlap between different samples. The techniques can be used for overlap control between samples for the same survey selected at different time periods or between different surveys at the same time period. Although the methods are in wide use, their properties, when a population is changing due to births and deaths, have not been studied extensively. Ideally, each technique should produce a sample proportionally allocated to births and persistent units when equal probability sampling is used. We present theoretical and empirical results showing the circumstances where proportional allocation is approximately obtained. We also discuss important cases where PRN and CRN sampling are substantially different in their coverage of birth and persistent units.

Ernst, Lawrence R. and **Ponikowski, Chester H.** (1998), **"**Selecting the Employment Cost Index Survey Sample as a Subsample of the National Compensation Survey," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 517-522.

A key aspect of the current integration of the BLS compensation surveys will be to select future establishment sample for the Employment Cost Index (ECI) survey as a subsample of the larger National Compensation Survey (NCS). Initially the transformation of the ECI to a subsample of the NCS will take place by initiating five subsamples of the current NCS sample over three years. This paper details our plans for the subsampling, which involves several interesting issues. For example, the NCS sample is actually two independent samples with some establishments selected twice, but the subsampling method insures that no establishment will be selected more than once in ECI. Also, the subsampling probabilities are determined with the goal that all establishments in an industry stratum in the universe with the same frame employment have the same overall chance of being in the ECI sample. Finally, we discuss the additional sampling complexities arising from the fact that the current NCS sample consists only of establishments with frame employment of at least 50, while the ECI has no

minimum employment restrictions, necessitating supplemental ECI samples of the smaller establishments.

Ernst, Lawrence R. (1998), "Maximizing And Minimizing Overlap When Selecting A Large Number Of Units Per Stratum With Simultaneous Selection," *Journal of Official Statistics,* 14, 297-314. Also in *Proceedings of the Section on Survey Research Methods*, American Statistical Association (1997), 475-480.

A number of procedures have been developed, beginning with the work of Keyfitz, for maximizing or minimizing the overlap of sampling units for two stratified designs. Most of these procedures are not applicable at all, or are not feasible to implement, unless the number of units selected per stratum is very small. The author previously presented a procedure for increasing or decreasing overlap when a large number of units per stratum are selected and when the sample units for the two designs must be selected sequentially, as is the case when the second design is a redesign of the first design. That procedure does not generally yield an optimal overlap. In the current paper a very different large sample per stratum procedure is presented for maximizing or minimizing overlap when the units can be selected for the two designs simultaneously, as may be the case for two different surveys. The procedure guarantees an optimal overlap if the two designs have identical stratifications, but can still be used, with loss of optimality, if the stratifications differ. An application of this procedure to the joint selection of samples for two BLS compensation surveys is discussed.

Tourangeau, Roger, Shapiro, Gary, Kearney, Anne and **Ernst, Lawrence** (1997), "Who lives Here? Survey Undercoverage and Household Roster Questions," *Journal of Official Statistics*, 13, 1-18.

We carried out an experimental comparison of three questions for enumerating residents of a dwelling. One version took the approach used in many surveys; this version began by asking respondents to name all persons living at the dwelling. The experimental versions began by asking how many persons had spent the previous night at the dwelling and used other probes to complete the roster. The two experimental versions differed only in that one version did not require persons to be listed by their full names, asking respondents to use initials or nicknames instead. A total of 509 interviews were completed, about a third with each version of the questionnaire. The results indicated that both experimental versions of the roster questions yielded more persons per household than the standard version; however, only the version that did not require full names yielded more persons identified as usual residents of the dwelling. Additional

analyses indicated that the same types of persons were listed on all three versions; only the number of persons differed.

Weidman, Lynn and **Ernst, Lawrence R.** (1996), "Multiple Workloads per Stratum Designs," *Journal of Official Statistics*, 12, 365-384.

This paper introduces an approach to expanding a stratified design, D_1, with one primary sampling unit (PSU) selected per stratum to a larger design, D_2. Define a workload (WL) to be the sample size in a given stratum in D_1. The three-stage approach selects the number of WLs for each stratum, the PSUs to receive additional WLs in each stratum, and the ultimate sampling units. Procedures are given for selecting PSUs in the key second stage, satisfying the following conditions when a stratum in D_2 is to have $s \geq 2$ WLs: (i) the expected number of WLs in a PSU is s times the probability that it was selected to get the single WL in D_1; and (ii) the actual number of WLs assigned is within one of the expected number. These conditions are a generalization of probability proportional to size, without replacement sampling. The properties and variances of this approach are compared to those from three alternative expansion procedures via application to a proposed, but since canceled, expansion of the Current Population Survey.

Ernst, **Lawrence R**. (1996), "Maximizing the Overlap of Sample Units for Two Designs with Simultaneous Selection," *Journal of Official Statistics,* 12, 33-45.

It is demonstrated, using transportation theory, that controlled selection can be used to solve the following sampling problem. Sample units are to be selected with probability proportional to size for two designs, both one unit per stratum, denoted as D1 and D2, with generally different stratifications. The goal of the problem is to simultaneously select the sample units for the two designs in a manner which maximizes the expected number of units that are in both samples. The procedure differs from previous overlap procedures in that it yields a better overlap, but is only applicable when the two samples can be selected simultaneously. An important special case occurs when the probability of selection for each unit in D1 does not exceed its probability of selection in D2. The procedure can then guarantee that the D1 sample units are a subset of the D2 sample units. A proposed, but since canceled, expansion of the Current Population Survey, which is discussed, would have been a potential application of this special case. Variance formulas for estimators of total under the controlled selection procedure are also presented. In addition, it

is demonstrated that the procedure can easily be modified to minimize expected overlap instead of maximizing it.

Esposito, James L., (1999) "Evaluating the Displaced Worker/Job-Tenure Supplement to the CPS: An Illustration of Multimethod Quality Assessment Research," Paper Presented at the Conference of the Federal Committee on Statistical Methodology.

Displaced workers are persons who have lost or left jobs involuntarily as a result of unfavorable economic events/conditions (e.g., plant closings, insufficient work, downsizing). The main objective of the paper is to illustrate the utility of using multiple methods to evaluate survey questionnaires. Three methods were used: (1) interaction coding (via behavior coding); (2) interviewer debriefing (using a focus-group format and rating forms); and (3) respondent debriefing (using follow-up probe questions). An effort is made to demonstrate how the qualitative and quantitative information provided by these techniques contributed to our "understanding" of how data quality may have been affected by various sources of measurement error.

Esposito, James L. and **Fisher, Sylvia** (1998), "A Summary of Quality-Assessment Research Conducted on the 1996 Displaced-Worker/Job-Tenure/Occupational-Mobility Supplement," *BLS Statistical Notes* (No. 43), Bureau of Labor Statistics, Washington, DC.

This paper summarizes research undertaken as part of the bureau's ongoing effort to improve the quality of its surveys. The DW/JT/OM supplement provides data on several issues that have been the focus of much research in recent years, including job loss due to structural change in the economy and the extent of employment instability. Three survey evaluation methods were used in conducting this research: behavior coding, interviewer debriefing using a focus-group format, and respondent debriefing using follow-up probe questions. As with any survey attempting to measure complex labor market phenomena, the quality of the information obtained from the DW/JT/OM supplement depends on clear definitions of the concepts to be measured and well-designed survey questions. The original goal of the quality assessment research was to examine the quality of the survey questions. What became apparent early in the evaluation process, however, was that there were many unresolved issues with the concepts being measured. To resolve these conceptual issues, a multiphase research plan is proposed for evaluating and possibly redesigning the DW/JT/OM supplement.

Fisher, Sylvia K. (2000), "Improving the Quality of Data Reporting in Business Surveys:

Discussant Comments," *Proceedings of the International Conference on Establishment Surveys II*.

Discussant comments focus on papers devoted to improving data reporting in business surveys. Willimack, Sudman et al. posit a model delineating what transpires in large business establishments at each stage of the response process. Eldridge et al. describe cognitive interview procedures conducted by the Office for National Statistics, United Kingdom, to improve business surveys. Babyak et al. map out procedures implemented by Statistic Canada's to pretest their establishment questionnaires. Both papers prescribe useful strategies to ameliorate the effects of barriers to conducting cognitive interviews in establishments. All three papers demonstrate the efficacy of cognitive methods in improving establishment data reporting.

Fisher, Sylvia. K., Ramirez, Carl, Stanley McCarthy, Jaki, and Shimizu, Iris**,** (2000), "Examining Standardization of Response Rate Measures in Establishment Surveys " *Proceedings of the Council of Professional and Federal Statistics.*

The Interagency Group on Establishment Nonresponse (IGEN) has studied issues associated with unit nonresponse in Federal establishment surveys and published proposals for research in this area. Previous work has documented the types of response rates calculated across a number of establishment surveys conducted by U.S. statistical agencies. In this paper: 1) we review the similarities and differences among those rates; 2) discuss the extent to which there is standardization in the calculation and publication of such rates and the reasons why this is so; and, 3) explore the possibilities, advantages, and disadvantages of fostering greater coordination and standardization across IGEN-member agencies in this regard.

Fisher, S. K. and **Kaspryzk, D.** (1996), "Revising the NCES Private School Survey: A Method to Design A Systematic Classification of Private Schools in the United States,". Paper presented at the American Statistical Association, Chicago, Illinois.

The Bureau of Labor Statistics (BLS) has been contracted by the U.S. Department of Education, National Center for Education Statistics (NCES) to conduct a series of cognitive laboratory and field tests on five specific questions taken from their national paper-and-pencil "Private School Survey (PSS)." The purpose of the PSS is to identify and classify private schools in the United States for federal-level statistical reporting. Administered by mail and/or telephone to approximately 27,000 U.S. private schools, results are used to classify U.S. private schools into one of nine categories that form a typology based on governance and program type. Study results were used to revise PSS items. Cognitive testing conducted by the BLS emphasizes developmental questioning and directed probing to identify subjects' response processes. Final testing focused on 1) pretesting revised PSS items; 2) a mode effect study contrasting mail-in

versus CATI administration; and 3) a validation study to estimate degree of respondent error with the new question(s).

Fisher, S. and **Stinson, L.** (1996), "Revising the NCES Private School Survey: A Method to Design a Systematic Classification of Private Schools in the United States," Paper presented at the American Educational Research Association, New York City, New York.

This study describes the efforts to validate items used to classify all private schools in the United States. The high number of participating private schools in the United States (nearly 27,000) makes the results of the study relevant to educators throughout the nation. This study provides an in-depth description of procedures used by the Federal Government and other institutions that conduct large-scale survey administration to analyze survey items for sources of measurement error, as well as the methodology used to eliminate as many sources of measurement error as possible.

Fisher, S. and **Stinson, L.** (1996), "Issues in Cleaning Up Your Data: The Example of the Bureau of Labor Statistics," Paper presented at the Southern Association of Public Opinion Researchers, Raleigh, North Carolina. .

Researchers engaged in the administration of large-scale regionally- and nationally-administered surveys encounter numerous difficulties when processing incoming data. Many problems in data collection can affect and impair the quality of the data generated from these surveys. "Cleaning" is an important aspect of enhancing data quality for purposes of statistical analysis. Data editing encompasses this entire process of cleaning data to verify its accuracy and ensure its quality. The purpose of this study is to identify issues associated with cleaning survey data, using the example of establishment survey activities conducted by the Bureau of Labor Statistics (BLS). This paper will describe those data editing activities and processes currently implemented to clean the data collected by BLS. We also document similarities and differences in currently applied data editing procedures, describe both the common and unique decision rules used by the various BLS surveys, and identify shared editing processes used for both household and establishment surveys.

Harris-Kojetin, Brian A., and **Fricker, Scott** (1999), " The Influence of Environmental Characteristics on Survey Cooperation: A Comparison of Metropolitan Areas," Paper Presented at 28th Session International Conference on Survey Nonresponse– Portland, OR. .

A request for survey participation takes place within a broad context – a social and economic environment that can vary over time, across societies, or even across different geographic areas within a society. Over these diverse contexts, there may be differing norms and expectations for interacting with strangers and for complying with requests for help. In the context of solicitation for survey participation, the end result may be varying response rates across areas of a country. House and Wolf (1978), Groves and Couper (1998), and others, have demonstrated such variations, particularly between urban and rural areas. In this paper we begin by reviewing the literature from social psychology, sociology, and survey methodology that identifies likely ecological correlates of survey cooperation. Based on these findings, we operationalize indicators of the demographic, social, and economic environment, and build composite indices to reflect the putative social psychological attributes of metropolitan areas in the United States. We then examine the extent to which these indicators are related to differing levels of survey cooperation across areas. As a necessary first step in developing a fuller nonresponse model, we begin by looking at response rate data from just two large-scale, national surveys- one public and one private- the Current Population Survey (CPS) and the Arbitron Radio Diary survey. First, we examine the pattern of cooperation rates for the two surveys to see whether different surveys conducted by different organizations experience similar relative cooperation rates in specific metropolitan areas. Secondly, we construct regression models utilizing response rates from both surveys as dependent variables, and the environmental indicators of the metropolitan areas as predictors. Finally, we discuss the implications of these findings for theories of survey cooperation and for improving data collection procedures.

Kominski, Robert, **Frazis, Harley,** Kaye, Kelleen, and McArthur, Edith, (2000), "How do we know what we've got? Looking at Measures of Educational Attainment across the Federal Government," *Proceedings of the Section on Social Statistics,* American Statistical Association.

The Interagency Committee on Measures of Educational Attainment is charged with proposing a standard set of attainment categories. The committee has reviewed the various measures used by federal agencies in their data collections. It has arrived at a consensus on a set of attainment levels for general purpose uses and is proposing further research in an area that is of great interest across the federal agencies. This discussion will review the committee's findings, describe the range of policy needs for data and the proposed future research.

Fullerton, Howard, N. Jr. (1997), "Evaluating The 1995 Labor Force Projections," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

The Bureau of Labor Statistics prepares labor force and employment projections about every two years. Six labor force projections over the 1980 to 1991 period were to or for 1995. The six labor force projections to 1995 are reviewed for accuracy, using varied measures of accuracy. These measures reflect different loss functions of users. The accuracy of the labor force projections to 1995 is compared with the accuracy of previous projections. Some possible reasons for the divergence of the actual and the projected are considered.

Fuxa, Mary Lee (1998), "Estimating Price Relatives for the U.S. Consumer Price Index When Priced Items are Restratified," *Proceedings of the Section on Business and Economic Statistics*, American Statistical Association, 137-142

For the 1998 revision of the Consumer Price Index, the stratification of priced items was extensively redefined. Of the 183 item strata: 33% did not change; 24% were designated for new sampling; 43% changed but were not designated for new sampling due to budgetary constraints. Until new samples are rotated in for the 43%, an adjusted method of estimation is needed for the monthly price relatives used to compute the index. The price relative for an item stratum and geographic area is a ratio of the sum of weighted price quotes for the current month to the sum for the previous month. This paper presents a method for adjusting the quote-level weights. Estimates with adjusted weights are computed for a test geographic area using the retained sample. They are compared to estimates with nonadjusted weights that are computed using a newly selected sample.

Garner, Thesia, Stinson, Linda and **Shipp, Stephanie** (1996) "Subjective Assessments Of Economic Well-Being And Preliminary Findings From Miami" *Proceedings of Section on Survey Research Methods,* American Statistical Association.

Surveys have long been used to measure both objective and subjective phenomena. Objective realities, such as one's date of birth or educational attainment, are matters of public record. Subjective reality, on the other hand, is totally private, it can be reported only by the person experiencing it. In this study we are concerned with respondents' private views of their income. These views may be included in their subjective assessments of income and may help to determine whether or not they experience some sense of economic well-being. Since most American adults are confronted

daily with the delicate task of balancing their income and expenses, it seems reasonable to expect them to have a rather well-developed budgetary sense.

Gentle, James E., Narula, Subhash**,** and **Valliant, Richard** (1996), "Multicriteria Optimization in Sampling Design," in *Statistics of Quality: Dedicated to Don Owen*, eds. S. Ghosh, W. Schucany, and T. Smith, New York: Marcel Dekker, 411-425.

Many problems in statistics and quality improvement are optimization problems. To determine an optimal sample design is a functional maximization problem in which the design points are chosen to maixmize information. In process improvement, the objective is to optimize some measure of expected performance over a range of process parameters. This paper reviews multicriteria optimization and describes a general sample optimization problem and software developed to solve the problem. Other applications of multicriteria optimization to robust regression estimation and nonparametric density estimation are described.

Getz, Pat, Kropf, Jurgen and **Strifas, Sharon** (1997), "Measuring The Contribution Of Business Births And Deaths To Overall Employment Movements," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 285-288.

One of the nation's most widely watched economic indicators is the monthly estimate of nonfarm payroll employment from the Current Employment Statistics program, a sample survey of nearly 400,000 business establishments. Because of the dynamic nature of the U.S. economy, business births and deaths can be significant contributors to the overall month-to-month movements in nonfarm employment, yet the measurement of employment change resulting from these components has always been problematic. This difficulty is largely attributable to the complications of maintaining an up-to-date sampling frame for a continually changing universe of business establishments, within a monthly survey environment. Improved universe information now provides an opportunity to comprehensively profile the contributions of birth, death, and continuing business establishments to employment movements. In this paper, profiles of the employment change contributions from these three basic components are developed. Seasonal and cyclical patterns emerge to form the basis for testing time series modeling techniques to estimate a net business birth/death component of monthly employment change, as an alternative to monthly sample frame maintenance and supplementation. The paper reports results from testing the model-fitting and forecasting accuracy of regression, ARIMA, and exponential smoothing techniques.

This study is undertaken within the broader context of a comprehensive program of methodological improvement for the CES survey.

Goldenberg, Karen, and **Phillips, May Anne,** (2000) "Now that the Study is Over, What have You Told Us? Identifying and Correcting Measurement Error in the Job Openings and Labor Turnover Survey Pilot Test," *Paper Presented at the International Conference on Establishment Surveys II, Buffalo, NY, June 2000 (Proceedings Forthcoming).*

This paper describes qualitative research by the Bureau of Labor Statistics (BLS) to identify sources of measurement error in the Job Openings and Labor Turnover Survey Pilot Test. Over the year of pilot data collection, BLS refined the concepts being studied and modified the data collection form. BLS subsequently conducted in-depth debriefing interviews with a small group of Pilot Test participants and compared their data with intended survey concepts. The results of these interviews were used to identify and correct additional sources of measurement error, and to further refine the data collection forms.

Goldenberg, Karen L., and **Stewart, Jay** (1999), "Earnings Concepts and Data Availability for the Current Employment Statistics Survey: Findings from Cognitive Interviews," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The Current Employment Statistics (CES) survey is conducted monthly by the Bureau of Labor Statistics (BLS) for the purpose of estimating month-to-month changes in total payroll employment, and in payroll earnings and hours for production and nonsupervisory workers. BLS is exploring the feasibility of using an earnings concept that is more inclusive than payroll, and collecting earnings and hours data for all employees. In a series of in-depth interviews with CES respondents, we explored respondent understanding of key concepts and survey terminology. We debriefed respondents on their current reporting practices and inquired about their information systems. We found that respondents had some ambiguity in their understanding of our basic concepts. In addition, we asked respondents whether they could provide the proposed items. Although almost all respondents answered affirmatively, further inquiries suggested otherwise. By expanding "cognitive" interviews to include detailed information about data systems relative to specific concepts, we learned about data quality and potentially avoidable sources of measurement error. These findings will be used to better understand the results of a data collection pilot study.

Goldenberg, **Karen L.,** Levin, Kerry, Hagerty, Tracey, Shen, Ted and Cantor, David (1997), "Procedures for Reducing Measurement Error in Establishment Surveys," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 994-999.

Measurement error in establishment surveys comes from the same sources as household surveys, but also derives from sources specific to establishments. This paper describes procedures used to reduce measurement error contributed by the respondent, the questionnaire, and the interviewer in the Business Births Pilot Study, an establishment survey which evaluated the feasibility of using CATI interviews to identify new businesses and obtain employment counts from them. Questionnaire testing methods included: (1) cognitive pretesting to assess respondent understanding of key concepts; (2) iterative telephone paper-and-pencil pretesting; and (3) behavior coding. Contact procedures consisted of (1) telephone screening interviews to locate the business and identify the most knowledgeable respondent, and (2) personally addressed advance letters, faxed or mailed to the named respondent. Questionnaire revisions using these procedures resulted in a smooth, easily administered, standardized telephone interview. Prescreening and advance letters effectively identified the correct respondent and encouraged participation. Eighty-five percent of completed interviews were conducted with the initially-designated respondent, and overall response rates averaged 82% per month among successfully-located businesses.

Goldenberg, Karen L. (1996), "Using Cognitive Testing in the Design of a Business Survey Questionnaire," *Proceedings of Section on Survey Research Methods,* American Statistical Association, 944-949.

Bureau of Labor Statistics' (BLS) staff used cognitive testing to evaluate and refine the questionnaire for the Current Employment Statistics (CES) Business Births Pilot Study. The CES survey is the source of current monthly U.S. payroll employment data. At present, it does not directly measure employment from newly-created businesses, or "business births." The means of selection into the CES sample is through an employer's Unemployment Insurance (UI) account number, and there is a lag of as much as 8 months between the time an employer obtains the account and the time it appears in the sampling frame. To collect data from business births near the time of their inception, BLS developed a procedure to sample businesses shortly after they obtain new UI accounts. The Business Births Pilot Study, getting underway in 1996, is a telephone survey designed to test the feasibility of differentiating new businesses from ongoing businesses that are undergoing changes such as new ownership,

incorporation, or merger. The telephone interview will determine whether a business is a birth, and if so, will obtain employment data and an industry description for it. The paper reports on the types of questionnaire problems identified through cognitive interviews, and on serendipitous findings about respondents' understanding of basic survey concepts. .

Horrigan, Michael and Herz, Diane (1999), "A Study in the Process of Planning, Designing and Managing a Survey Program: The case of time-use surveys at the BLS," *Proceedings of the Section on Social Statistics,* American Statistical Association.

The BLS has recently set up a working group for the purpose of examining the possibility of conducting a time-use survey. This examination follows two significant activities on time use that the BLS has undertaken in the last year: (1) A pilot study of two alternative versions of a time-use survey using a telephone methodology, and (2) co-sponsorship (with the MacArthur Foundation) of the November 1997 conference on time use. Specifically, the working group has been charged with developing a report that addresses the feasibility of conducting a time-use survey using a subsample of the outgoing rotation groups from the monthly Current Population Survey. If viewed as technically feasible, the report will also develop a comprehensive management plan for the design and execution of such a survey. The working group is developing a report that will be delivered to Commissioner Abraham during the summer of 1998. The paper for the ASA session will provide an insider's view to the process of developing a survey strategy as well as a summary of the working group's recommended approach.

Wolter, Kirk, Shao, Jun and **Huff**, **Larry** (1998), "Variance Estimation for the Current Employment Statistics Survey," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 775-780.

The Current Employment Statistics (CES) Survey supplies some of America's most vital leading-economic indicators. At the 1997 annual meeting in Anaheim, various authors reported on the status of an ongoing redesign of this important survey. Their papers discussed the rationale for the redesign, the new sampling design, the new estimation procedures for employment levels and trends, and special estimation issues for hours and earnings variables. In the current paper, we discuss variance estimation for all CES variables. The overall variance estimation strategy involves primary use of the balanced half-samples (BHS) method for higher-level, aggregate statistics, and secondary use of generalized variance functions (GVF) for the more disaggregated statistics. Our BHS method addresses many CES

design features, such as stratification, more than 2 primary units per stratum, clustering, birth and death sampling, and rotation sampling. We incorporate within our method allowances for the imputation variance and for the finite population correction. And, to improve variance estimation for nonlinear statistics, we employ both the half sample estimator and its complement. In addition to describing the variance estimators, we present the results of a simulation study done to shed light on the statistical properties of the estimators.

Jacobson , Shawn (1999) *"*Relationship Between Data Quality and Collection Date in the Consumer Price Index Housing Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

In order to estimate shelter cost inflation, BLS collects Rent data from a sample of renters for the Consumer Price Index. Each rental unit is interviewed every six months; in each interview, the respondent is asked what the rent is in the current month and what the rent was in the previous month. One-month rent relatives are calculated using these reported rent values; six-month rent relatives are computed using current month rent data from preceding interviews. This paper gives the results of an exploratory data analysis based on rent data from 1992 through 1996. The results indicate that the proportion of six-month rent changes increases and the proportion of one-month rent changes decreases as the collection date increases throughout the month. Variances are computed using the stratified random-groups method, to test the significance of these trends. More recent data (from 1997 and 1998) is used to determine if these results hold in later time periods. Possible (psychological) explanations of the phenomena are considered.

Jacobson, Shawn, Leaver, Sylvia and **Swanson, David** (1998), "Choosing a Variance Computation Method for the Revised Consumer Price Index," *Proceedings of the Section on Business and Economics Statistics*, American Statistical Association 131-136.

In this paper the authors compare six candidate variance methods three hybrid linearization methods, balanced repeated replication, unstratified jackknifing, and stratified random groups, for twelve published CPI index series. The stability of the resulting variance series was used as a proxy for the variance of variance estimates. Three non-numerical criteria (cost, ease of implementation, and speed) were also evaluated. No candidate method was most stable for all examined index series, and no method was best at meeting all criteria; however, the stratified random groups method

(implemented using VPLX) was marginally better than the other candidate methods.

Jain, Raj (1998), "Measurement Errors and State Space Model Based Method of Seasonal Adjustment: Some Empirical Results," *Proceedings of the Section on Business and Economics*, American Statistical Association.

The data for various unemployment series published by the Bureau of Labor Statistics (BLS) are collected using a rotating sample design. The effect of such a sample design is to make the sampling errors in the data series, serially correlated. Most seasonal adjustment methods do not take account of the correlated structure of such errors; hence, these methods produce less efficient estimates of the trend and seasonally adjusted series. In the SSMB method of seasonal adjustment, an unobserved component model for these measurement errors is introduced. The model incorporates an ARMA process for these errors. Two unemployment rate series, published by the BLS, are seasonally adjusted by the SSMB method, and the effects of measurement errors are analyzed.

Jain, Raj K. (1997), "Automatic Outlier Detection And Adjustment In Seasonal Adjustment Methods," *Proceedings of the Section onBusiness and Economics Statistics*, American Statistical Association., 217-222

Several BLS price index series such as gasoline price and women's apparel price series are affected by non-economic interventions which generate outliers in the series. To seasonally adjust these series account has to be taken of these outliers. The automatic outlier detection and adjustment of the series by three seasonal adjustment methods, the State Space Model Based Method (SSMB), Bureau of Census's X-12ARIMA method and Burman's ARIMA model based method, is studied. The results of estimation by the three methods are then compared with respectto the quality of seasonal adjustment of these series.

Jain, Raj K. (1996), "Trading Day And Easter Adjustment In Seasonal Adjustment Methods" *Proceedings of the Section on Survey Research,* American Statistical Association.

Several retail sales series which are affected by the number of trading days in a month and the moving nature of the Easter, are seasonally adjusted by the State Space Model Based (SSMB)method , the Bureau of Census's X-12 ARIMA method and Burman's ARIMA model based method of

seasonal adjustment. The results of estimation from these methods are then compared with respect to the quality of seasonal adjustment.

Jayasuriya, Bodhini R. (1996), "Testing for Polynomial Regression Using Nonparametric Regression Techniques," *Journal of the American Statistical Association*, Vol. 91, pages 1626-1631.

In regression analysis, it is important to test the validity of the assumed model prior to making inferences regarding the population of interest. In this investigation, we utilize nonparametric regression techniques to test the validity of a kth order polynomial regression model. The departures from the polynomial model are assumed to belong to a smooth class of functions; a parametric form is not assumed. A test based on nonparametric regression fits to the residuals from k^{th} order polynomial regression is proposed. It utilizes a smoothing spline fit of order $2k$ to the residuals from k^{th} order polynomial regression. A test statistic based on this estimator is formulated and its asymptotic distribution is derived under alternatives converging to the null at a rate of $\left(n\mathbf{l}^{1/4k}\right)$ where $\mathbf{l}$ is the smoothing parameter. We note that this rate of convergence is slower than the parametric rate of $n^{-1/2}$. Power investigations are conducted through a small scale simulation study.

Jayasuriya, Bodhini R. and **Valliant, Richard** (1996), "An Application of Restricted Regression Estimation in a Household Survey," *Survey Methodology,* Vol. 22.

This paper empirically compares three estimation methods—regression, restricted regression, and principal person—used in a household survey of consumer expenditures. The three methods are applied to post-stratification which is important in many household surveys to adjust for under-coverage of the target population. Post-stratum population counts are typically available from an external census for numbers of persons but not for numbers of households. If household estimates are needed, a single weight must be assigned to each household while using the person counts for post-stratification. This is easily accomplished with regression estimators of totals or means by using person counts in each household's auxiliary data. Restricted regression estimation, refines the weights by controlling extremes and can produce estimators with lower variance than Horvitz-Thompson estimators while still adhering to the population controls. The regression methods also allow controls to be used for both person-level and household level counts and quantitative auxiliaries. With the principal person method, persons are classified into post-strata and person weights are ratio adjusted to achieve population control totals. This leads to each person in a household potentially having a different weight.

The weight associated with the "principal person" is then selected as the household weight. We will compare estimated means from the three methods and their estimated standard errors for a number of expenditures from the Consumer Expenditure survey sponsored by the Bureau of Labor Statistics.

Johnson, William, (2000), "Exploratory use of Spatial Statistics to Inform the Choice of Variables for Stratifying Non-Self Representing Cities for the Next CPI PSU Sample Selection," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

After data from the 2000 Census becomes available, all non-self representing areas will be stratified prior to selecting the PSU sample for the Consumer Price Index (CPI). This paper discusses the findings from the use of spatial statistics and exploratory spatial data analysis to examine the presence of spatial correlation in 12-month price change for the all items CPI-U during the time period from 1988 through 1999. The models of 12-month price change considered in searching for stratifying variables for the sample design are examined for the presence of spatial lag and spatial error during this period. Earlier work done within the Bureau of Labor Statistics, which led to the selection of stratifying variables based on these models, is discussed in light of what is found.

Johnson, William, **Leaver, Sylvia G.,** and **Benson, Thomas S.** (1999) "Modeling the Realized Outlet Sample for the Commodities and Services Component of the U.S. Consumer Price Index," *Proceedings of the Section on Government Statistics,* American Statistical Association.

This paper describes the use of simulations to model the realized outlet sample for the commodities and services component of the U.S. Consumer Price Index as part of the effort to optimize the sample design. Outlet samples for different categories are drawn from independent frames obtained through the Telephone Point of Purchase survey. Outlets may occur in multiple categories or half samples, so the realized sample may have fewer unique outlets than the initial sample size specified. The efforts to model the actual number of unique outlets realized in spite of difficulties arising from the use of several sources of data are documented, and an attempt is made to evaluate the effectiveness of the resulting models. As the sample composition is changing over time from one drawn entirely from frames obtained through the Consumer Point of Purchase Survey to one drawn entirely from frames obtained through the Telephone Point of

Purchase Survey an attempt is made to assess how the introduction of frames from this new source affects the realized sample.

Kojetin, Brian A. and Tanur, Judith M. (1996), "Youths Proxied And Proxying: Communication And Reports Of Job Search" *Proceedings of the Section on Survey Research,* American Statistical Association, 254-259.

A program of research has examined the reporting of job searches by youths and the adults who most often proxy for them in the Current Population Survey (CPS), and the possible impact on the youth's unemployment rate. A better understanding of this differential reporting has been hampered by a lack of basic knowledge of intra-household communication patterns. Recently, research conducted by the Bureau of Labor Statistics (BLS) has focused on communication within the family about issues on which proxies report in the CPS. All family members at least 16 years of age from 97 households participated in a study in which they answered CPS questions about themselves and each other and reported how they learned about the other's activities. This design eliminates the usual self-selection into self-reporting vs. proxy reporting status, as each subject acts both as a self-reporter and as a proxy, enabling self and proxy responses to be compared. In this paper, we will compare the communication between young people and their proxies with communication between older people and their proxies and will further investigate whether any differences found in these communication patterns are consequential for reports of job search on the CPS.

Kratzke, Diem Tran, Shierholz, Heidi S. and **Woodruff, Stephen** (1997), "Measuring Employment From Births and Deaths in the Current Employment Statistics Survey," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 210-215.

The Bureau of Labor Statistics is redesigning the Current Employment Statistics survey, which is a survey of about 390,000 businesses that is the source of current monthly U.S. payroll employment statistics. A major concern with the redesign is the reliable measurement of employment from new businesses (births) in a cost effective and timely manner. One method for measuring birth employment is to measure employment lost from business deaths (death employment) and add that to a measure of net employment, where net employment is the difference between birth employment and death employment. The goal of our research was to develop statistical models to predict net employment. Monthly establishment employment for many years and across industries was available from the Business Establishment List for the research. Some of

the statistical tools used were seemingly unrelated regression and mixed estimation. Theory and empirical results are presented and evaluated.

Kropf, Jurgen , Getz, Patricia, and **Roosma , Michael** (1999), *"C*auses and Treatments of Non-Economic Fluctuations in Average Weekly Hours and Average Hourly Earnings Series*," Proceedings of the Section on Government Statistics,* American Statistical Association.

The Current Employment Statistics (CES) Survey is conducted monthly by the Bureau of Labor Statistics (BLS) for the purpose of estimating over-the-month changes in total employment, and in payroll earnings and hours for production and non-supervisory workers. In 1997, researchers noted the presence of fluctuations in the Average Weekly Hours (AWH) and Average Hourly Earnings (AHE) series, that appeared to be non-economic in nature and related to variations in the calendar. In this paper, we discuss the process used to identify and treat the observed distortions. First, the microdata were screened by an equal means test to indicate significant differences in months with varying numbers of days per pay period. These tests identified problematic industries and reporters. Second, we generated estimates without the problematic reports; this resulted in near elimination of the fluctuations in AWH and mitigated the distortions in AHE. Next, we contacted reporters to inquire about their reporting practices and confirmed the source of variation in the payroll figures. Finally, we developed time series models with variables designed to identify, measure, and treat the effects of the varying length of pay periods. The modeling resulted in the successful treatment of affected industries, which now display seasonal adjusted series with virtually no calendar-related fluctuations and significantly increased smoothness.

Kydoniefs, Leda and **Stinson, Linda** (1999) "Standing on the Outside, Looking In: Tapping Data Users to Compare and Review Surveys," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

It is not unusual for survey organizations to conduct periodic expert review (sample design, definitions, question wording, data collection procedures) of ongoing survey programs. But the question arises, is there another way to conduct such a review? What if attention was turned toward the individuals who use the data? This paper presents the methodological approaches one could consider in pursuing answers from users to the questions implied by "what do data users want?" It also considers which approach provides us with the most extensive information about survey users given limited time and resources.

Leaver, Sylvia G., Johnson,William, Shoemaker, Owen and **Benson, Thomas S.** (1999) "Sample Redesign for the Introduction of the Telephone Point of Purchase Survey Frames In the Commodities and Services Component of the U.S. Consumer Price Index ," *Proceedings of the Section on Government Statistics,* American Statistical Association.

This paper describes the methodology for the redesign of the sample for the commodity and services component of the U.S. Consumer Price Index to accommodate the introduction of outlet frames from the Telephone Point of Purchase Survey. This work represents a further expansion and revision of models developed for the 1998 CPI sample redesign. Models relating data collection costs and sampling variance of price change to item and outlet selection variables for the sample design were developed and estimated. With these models, data collection resources were allocated to minimize sampling variance of price change, subject to budgetary and operational constraints, using nonlinear programming techniques. Models for sampling variance and costs are given, and solutions to the design problem posed under varying assumptions are discussed

Leaver, Sylvia G. and **Cage, Robert A.** (1997), "Estimating The Sampling Variance For Alternative Estimators Of The U.S. Consumer Price Index," *Proceedings of the Section on Government Statistics*, American Statistical Association.

The current official U.S. Consumer Price Index is a modified Laspeyres index. However, BLS has recently examined a number of alternative index series. A set of "superlative" index estimators, based on alternative methods of aggregating basic-level Laspeyres indexes, were published in December 1993, and a geometric mean index, based on a Laspeyres-type aggregation of geometrically averaged basic indexes, was published in April 1997. Sampling variances for the official Laspeyres formula index and for the geometric means index for shelter have been previously estimated and published. In this work, the sampling variances for the geometric mean index for non-shelter items and for superlative index estimators, computed using a stratified jackknife method, are studied. Additionally, contrasts between alternative superlative index estimators, and between Laspeyres and geometric mean index estimators and their variances are presented. The paper also compares the stratified jackknife variance estimator with the stratified random group variance estimator for the official Laspeyres and geometric mean shelter indexes, series for which implementation of both methods is possible.

.

Leaver, Sylvia, Johnson, William, Baskin, Robert, Scarlett, Samuel and **Morse, Robert** (1996), "Cost-Variance Modeling For The 1998 CPI Commodities And Services Sample Redesign," *Proceedings of the Section on Survey Research,* American Statistical Association, 239-244.

This paper describes the methods used to allocate data collection resources for the 1998 redesign of the sample for the commodity and services component of the U.S. Consumer Price Index. These methods rely on models relating data collection costs and sampling variance of price change to item and outlet selection variables for the sample design. With these models, the optimal allocation of data collection resources to minimize sampling variance of price change, subject to budgetary and operational constraints, can be found using nonlinear programming techniques. This work represents an expansion of models developed for the 1987 CPI sample redesign. Models for sampling variance and costs are given, and solutions to the design problem posed under varying assumptions are discussed.

Lee, Tae-Hwy and **Scott, Stuart** (1996), "Transmission of Producer Prices Through Stages of Processing," *Proceedings of the Section on Survey Research*, American Statistical Association, 110-119.

An economy can be subdivided sequentially into several stages, and producer price indexes (PPI) can be developed for each stage. The U. S. Bureau of Labor Statistics has developed two such formulations: one emphasizing the degree of a product's fabrication and another the direction of transactions among stages. Tracing the inflation transmission in terms of timing, direction, and magnitude is then possible. Inflation transmission is examined and compared in the two PPI systems using vector error correction models (VECM). The relationships between the long-run and short-run factors are explored to learn the dynamic structure of each price system. Forecasts of CPI inflation using the two PPI systems are also evaluated.

Lent, Janice, (2000). "Chain Drift in Some Price Index Estimators," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Since the market of available consumer goods is constantly changing, short-term price indexes may be directly measured more accurately than long-term indexes. Thus chained price indexes, which are essentially products of short-term indexes, are often used for estimating long-term changes in the cost of living. Differences between chained and direct indexes have been studied by Forsyth and Fowler (1981), Szulc (1983),

and others. Chained arithmetic mean indexes, such as the Laspeyres and Sauerbeck indexes, are known to be vulnerable to upward bias relative to their direct counterparts. This bias due to chaining is known as "chain drift." In this paper, we use analytical and empirical methods (involving data from the US CPI) to examine the effect of chaining on some geometric mean estimators. We find that, for a straightforward Tornqvist index estimator, the effect of chaining is unimportant under reasonable assumptions.

Lent, Janice, Miller, Stephen, Duff, Martha and **Cantwell, Patrick (1998),** "Comparing Current Population Survey Estimates Computed Using Different Composite Estimators," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 564-569.

We examine the effect of a new composite estimation method on estimates from the Current Population Survey, the U.S. labor force survey. Before December 1997, the *AK* composite estimator was applied directly for each characteristic of interest. To ensure consistency, the same coefficients were used for all estimates, though they were selected to be optimal only for unemployment totals. The new "composite weighting" method provides increased operational simplicity for micro data users and allows optimization of compositing coefficients for different labor force categories. Using both additive and multiplicative models, we develop adjustment factors that allow comparison of labor force estimates computed using the old and new composite estimators. The models appear to fit the data well; we illustrate their use for some key CPS statistics.

Lent, Janice, **Miller, Stephen** and Cantwell, Patrick, J. (1996), "Effect Of Composite Weighting On Some Estimates From The Current Population Survey" *Proceedings of the Section on Survey Research,* American Statistical Association, 130-139.

In a previous paper(Lent, Miller, Cantwell 1994), we considered methods of computing microdata weights for the Current Population Survey (CPS) that incorporate the effect of composite estimation for major labor force categories and demographic groups. These weights would simplify computation of composite estimates and improve the reliability of some important labor force estimates. In this paper we examine the effect of the composite weighting procedure on CPS estimates other than those for major labor force categories. We consider, for example, the reliability of estimates of numbers of (1) full- and part-time workers, (2) workers in various industry and occupation categories, and (3) persons unemployed by their reasons for unemployment. In addition, we present research on optimal parameter values for the AK composite estimator used in the CPS;

the optimal values may have changed with the redesign of the CPS implemented in 1994. We also discuss modifications to the composite weighting procedure necessitated by the recent CPS sample cut.

Levi, Michael (1997), "A Shaker Approach To Web Site Design," *Proceedings of the Section on Statistical Computing*, American Statistical Association.

As World Wide Web deployment (both over the Internet and through institutional intranets) continues to explode, developers are uncovering the myriad design possibilities and trade-offs that accompany this technology. Site designers must make decisions ranging from what content should be included, to the overall organization and sequence of pages, to placement and style of text, graphics, and buttons. As the Web matures, enhanced capabilities require additional decisions, such as when and how to replace static pages with CGI scripts and interactive Java or ActiveX applets. The statistical community, in particular, must address problems related to navigation through very large data sets and the accommodation of an expanded user population including persons less familiar with the subtleties of statistical analysis. Ultimately, the difference between a useful site and a frustrating one will be determined by the coherence of the developer's vision. In this session Michael Levi will draw on the Shaker ideals of simplicity, elegance, and quality to present a philosophy of Web site design based on the principles of Human-Computer Interaction and his view of Web site creation as a software development process. Mr. Levi will discuss his taxonomy of Web sites, address the importance of a user-centered design perspective, and present a set of usability principles tailored to Web systems. Mr. Levi will concentrate on the design task, and will not address the technical details of servers or development tools such as Perl or Java. This session is intended for Web developers and managers who wish to explore the issues of Internet or intranet development.

Mason, Charles, (2000), "Results from a Random Digit Dialing Survey of Where Consumers Purchase Goods and Services," *Proceedings of the Section on Government Statistics,* American Statistical Association.

The Bureau of Labor Statistics conducts a large household survey, the Point of Purchase Survey, which is used as the sampling frame for the selection of retail establishments priced in the computation of the Consumer Price Index (CPI). Within each of these outlets, BLS Field agents select unique items for repeated pricing from a category of goods and services independently selected from the expenditure data reported in the Consumer Expenditure Survey. In 1998, BLS converted the Point of Purchase Survey from a "personal visit" household interview survey based

on a decennial Census address sample to a "computer assisted telephone interview" survey using a list assisted random digit dialing sample. This paper describes the new survey and the flexibility it provides the BLS in updating our CPI outlets. In addition it reports on the first production data results from the new telephone based Point of Purchase Survey (TPOPS) and examines the response rates and data quality; especially the outlet address information, which is especially important in this survey since it is our only means of locating the outlets. In addition, comparisons of the number of retail outlet reported between TPOPS and the older POPS survey are examined.

McClelland, Robert and **Reinsdorf, Marshall** (1997), "Estimating Small Sample Bias In Two Price Index Formulas," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

BLS is able to estimate expenditures during the base period on individual items in particular stores, but a lack of information on base period prices prevents it from using these expenditures to find the quantity weights called for by the CPI's Laspeyres index formula. Using the later "link month" price both as the starting point for measuring price change and as a proxy for the unknown base period price leads to a positive covariance between errors in weighting and price changes. This effect has become known as "formula bias." Avoiding formula bias when estimating the price index for a particular item stratum in a particular area requires the use of a non-linear estimator whose expected value rises as the sample size falls. To get empirical evidence on formula bias and small sample bias, we draw samples from a simulated "population" formed by pooling CPI sample data from many areas. For long run indexes, the expected value of the geometric mean index is much more sensitive to sample size than the expected value of the "seasoned" index formula that BLS has adopted.

McConnell, Sheila, and **Goodman, William,** (1999), Recognition of More than Possible Trend in Time Series: Redesigned Screening of Microdata in the Current Employment Survey," *Proceedings of the Section on Government Statistics,* American Statistical Association.

In June of 1995, The Bureau of Labor Statistics (BLS) announced plans to redesign the method of selecting establishments to be included in the sample of the Current Employment Statistics (CES) survey. The newly designed sample is random, unlike the current sample. Partially in response to the recent sample design, new methods to check the quality of reported data are being developed. The new data editing and screening system will include basic checks for internal consistency of records and will also check data against a variety of plausible patterns in reported data over time. In the latter type of

test, an establishment's current reported values are compared to earlier data from the same establishment. The current values are screened using several tests. Passing just one of the tests results in acceptance of the data. This paper presents the editing and screening methods currently being tested and their performance

McKay, Ruth B. (1997), "The Multiracial Category As "Wild Card" In Racial Questionnaire Design," *Proceedings of the Section on Survey Research Methods (AAPOR)*, American Statistical Association, 916-921.

Previous analysis of data from the May, 1995 Current Population Survey CPS) Supplement on Race and Ethnicity had indicated that over 50% of the respondents who chose the multiracial category were "uncertain multiracials," i.e., did not identify multiple races, and that Hispanics tended to use the multiracial category as a substitute for "Hispanic" if that category was not available. In addition, the presence of the multiracial category led to a significant decline in population estimates for American Indians and Alaska Natives. Further analysis of the CPS Supplement data has added to the list of unanticipated outcomes of including a multiracial category. Although there is a decline in numbers of American Indians and Alaska Natives, very little of the shift is to the multiracial category itself. Rather, in the presence of the multiracial category, there is a greater tendency of members of the American Indian and Alaska Native group to identify as another race, e.g., white. In addition, only 31% of children of interracial married couples were reported as "multiracial."

McKay, Ruth B. (1996), "Cognitive Research to Reducing Nonsampling Errors," *Proceedings of STATISTICS CANADA Symposium '96,* Ottawa.

Cognitive research played an important role in reducing nonsampling errors in planning and implementing the CPS Supplement on Race and Ethnicity as well as evaluating the findings of the survey. Research interviews testing successive versions of the Supplement questions made it possible to reduce survey error by identifying and correcting problems in the instrument, including vague or imprecise questions, sensitive questions, abstract questions, vocabulary problems, order effects, and perceived redundancy. Monitoring and behavior coding CATI interviews, observing field interviews, and analyzing open-ended answers to CPS questions on "multiracal status," and "ancestry and ethnic origin," enabled us to identify non-sampling errors associated with interviewer behaviors. respondent difficulties, and problems unique to interviews as opposed to mail surveys. As the results of the CPS Supplement on Race and Ethnicity show, it will be difficult to eliminate nonsampling error in surveys of race and ethnicity until agreement is reached on the meanings assigned to the terms "race"

and "ethnicity/ethnic origin" by those who design surveys and those who respond to surveys.

McKay, Ruth B., **Stinson, Linda L**., de la Puente, Manuel, and **Kojetin, Brian A**. (1996), "Interpreting The Findings Of The Statistical Analysis of the CPS Supplement on Race and Ethnicity," *Proceedings of the Bureau of the Census' 1996 Annual research Conference,* Rosslyn, 326-337.

> This paper draws on the results of several cognitive research activities carried out in association with the CPS Supplement on Race and Ethnicity to improve the accuracy of interpretations drawn from statistical analysis of Supplement data.. Researchers conducted 83 cognitive interviews with respondents drawn from the major racial and ethnic groups in the U.S., monitored Supplement interviews in the Hagerstown and Tucson CATI facilities, conducted focus groups with interviewers in both CATI facilities, accompanied CPS field interviewers in Tucson and Miami, behavior coded four hundred CATI interviews (350 in English and fifty in Spanish), and analyzed the open-ended answers to CPS Supplement questions on "multiracial" status, and "ancestry and ethnic origin." The cognitive research findings enabled the researchers to classify "multiracial" persons on the CPS Supplement into "True Positives" and "False Positives," and explain the likely contributors to inaccurate multiracial reporting.

Mckay, R.B and de la Puente, M. (1996), "Cognitive testing of racial and ethnic questions for the CPS Supplement." *Monthly Labor Review*, 119, (9), 8-12.

> "Race" and "ethnic origin" are two of the most emotionally-charged and conceptually-ambiguous terms in contemporary social discourse. The task of developing the Current Population Survey (CPS) Supplement on Race and Ethnic Origin required extensive involvement of the behavioral sciences. An interagency team of behavioral scientists representing anthropology, psychology, and sociology, was closely involved in the development and testing of the questions for the May, 1995 CPS Supplement. The team's involvement continued during and after collection of the Supplement, assessing live and taped Supplement interviews, and contibuting behavioral science insights to interpretation of ambiguous findings from the statistical analysis of the data.

McKay, R., Breslow, M., and **Sangster, R. L.** et. al. (1996), "Translating Survey Questionnaires: Lessons Learned." *In New Directions in Program Evaluation Series: Current Issues in Survey Research*, Marc T. Braverman and Jana Kay Slater eds. New York: Jossey-Bass. pp.93-104.

The rapid increase in non-English speaking populations within the United States dictates the need for well-translated survey instruments to reduce nonresponse and measurement error. This chapter draws on three case studies where the authors dealt with a variety of experiences while managing the process of translating surveys into other languages (including Spanish, Mandarin \Cantonese, Khmer (Cambodian), Vietnamese, and Korean). It provides suggestions for designing better translations and discusses ways to avoid potential problems.

de la Puente, M. and **McKay, R**. (1996), "Cognitive Research is Making the Task of Asking Questions about Race and Ethnicity in Surveys a Little Easier," *Civil Rights Review*, vol. 1, no. 1, (In press).

Although race is a familiar concept, and discussions concerning this seemingly basic human characteristic are encountered on a daily basis in print and broadcast media, race is a difficult trait to measure in surveys. However, social scientists who study the cognitive aspects of surveys are providing valuable information on how individuals are providing valuable information on how individuals interpret and respond to survey questions, including questions on race and ethnicity. This article reports on one such effort, undertaken by a team of social scientists from the Federal and private sectors, to develop the Supplement on Race and Ethnicity for the May, 1995 Current Population Survey (CPS).

Mueller, Charlotte, and Phillips, Mary Ann, (2000), "The Genesis of an Establishment Survey: Research and Development for the Job Openings and Labor Turnover Survey at the Bureau of Labor Statistics," *Proceedings of the Section on Government Statistics,* American Statistical Association.

In late 2001, a new establishment survey will be making its debut at the Bureau of Labor Statistics (BLS). This survey--the Job Openings and Labor Turnover Survey (JOLTS)--will fill a current void in national economic data. For many years, the unemployment rate has served as an economic indicator of the unused labor supply. JOLTS intends to serve as a parallel indicator of unfilled labor demand. In the short time since 1999, JOLTS has been transformed from a concept to a live survey of 16,000 establishments across the United States. Monthly publication of rates and numbers of job openings, hires, and separations is the goal. This paper profiles the many research and development steps taken for creating this new survey. Looking outside BLS, we studied existing data sources similar in nature, and contacted academics for comments. Internally, we studied other BLS surveys and defined terms and data elements, established

reference periods, and designed forms and interviewer scripts. Developmental testing included a pilot test and reinterview surveys.

Paben, Steven P., (1999), " Comparison of Variance Estimation Methods for the National Compensation Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

A key use of the National Compensation Survey (NCS) is for the estimation of hourly mean wages for different localities. In NCS, hourly mean wages are calculated as the ratio of total annual pay to the total number of annual hours worked. In this study, an artificial Metropolitan Statistical Area (MSA) was created by combining NCS data from 16 different localities to serve as a sampling frame for 100 simulated samples. Then, the variance of the 100 sample estimates was compared to variance estimates obtained with the linearized Taylor Series method of variance estimation and three different methods of replication, balanced repeated replication (BRR), a Jackknife method, and Fay's method. To compare the variance estimates of the different methods a number of evaluative statistics were calculated, including bias, stability, and 95% confidence interval coverage.

Paben, Steven P. and **Ernst, Lawrence R.** (1997), "Adjusting Establishment Selection Probabilities And Number Of Occupational Selections To Reduce Variances In BLS Compensation Surveys," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 469-474.

The National Compensation Survey (NCS) is the new BLS survey for measuring employee wages by skill level. NCS uses a rotating panel design, with three stages of selection used in selecting each panel: geographic area PSUs; establishments selected from industry strata; and occupations selected separately from each sample establishment. The establishments are selected with probability proportional to size (pps), with total employment the measure of size. For each selected establishment the number of sampled occupations are a function of employment size. This paper investigates two aspects of this sample design. First, whether some other probability selection scheme for sampling establishments yields lower variances than direct pps. Second, if there exists an allocation by size class for the number of occupational selections that yields lower variances than the current allocation.

Paulin, Geoffrey *(1999), "*Applied Analyses of Multiply Imputed Income Data from the Consumer Expenditure Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

As is the case with many household surveys, income data collected in the Consumer Expenditure Survey are subject to item nonresponse. The current method for addressing this problem involves dividing participating families into "complete" and "incomplete" reporters. However, "complete" reporters need not provide a full accounting of income to qualify; in most cases, providing of data on one major source of income is sufficient. Multiple imputation has been proposed as a solution for item nonresponse to income questions in the Consumer Expenditure Survey. The Division of Consumer Expenditure Surveys (Bureau of Labor Statistics) has sponsored much research into the feasibility of developing a model-based imputation method to adjust for this problem. Experimental data have recently become available to internal researchers at the Bureau. How might these data be used in practice? And how, if at all, will results change when imputed data are used instead of "completely" reported data? This paper incorporates several types of analysis using these data and reports on the results.

Paulin, Geoffrey (1999) " Applied Analyses of Multiply Imputed Income Data from the Consumer Expenditure Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

As is the case with many household surveys, income data collected in the Consumer Expenditure Survey are subject to item nonresponse. The current method for addressing this problem involves dividing participating families into "complete" and "incomplete" reporters. However, "complete" reporters need not provide a full accounting of income to qualify; in most cases, providing of data on one major source of income is sufficient. Multiple imputation has been proposed as a solution for item nonresponse to income questions in the Consumer Expenditure Survey. The Division of Consumer Expenditure Surveys (Bureau of Labor Statistics) has sponsored much research into the feasibility of developing a model-based imputation method to adjust for this problem. Experimental data have recently become available to internal researchers at the Bureau. How might these data be used in practice? And how, if at all, will results change when imputed data are used instead of "completely" reported data? This paper incorporates several types of analysis using these data and reports on the results.

Paulin, Geoffrey and **Vannett, Valerie** (1999), *"*Enhancing the Product Line: Issues in Publishing Variances in the Consumer Expenditure Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The Division of Consumer Expenditure Surveys (DCES) is exploring new ways to improve data quality. One aspect of this has been the development of selected descriptive statistics for the average annual expenditures that are currently published. Tables have been developed that contain the mean expenditures, variance, standard error of the means, and coefficients of variation on an annual basis by demographic characteristics. The first section of this paper discusses the methods and issues surrounding the development of these descriptive statistics. The remaining portion of this paper details the uses of these statistics both within and outside the DCES. Within the DCES they will be used for quality control and as a research tool. At the same time, the public will have greater statistical information for the comparison of expenditures by demographic characteristics. Tables with descriptive statistics will be available upon request, and an example of a SAS program to produce these statistics will be added to the documentation for microdata available on CD-ROM.

Sverchkov, Michael, and **Pfeffermann, Danny,** (2000), " Prediction of Finite Population Totals Under Informative Sampling Utilizing the Sample Distribution," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The sample distribution is the distribution of the response variable for units included in the sample. This distribution is different from the population distribution if the sample selection probabilities are related to the values of the response variable. Previous studies of the use of this distribution focus on inference on the parameters and shape of the population distribution. In this article we study the use of the sample distribution for the prediction of finite population totals under single-stage sampling. It is assumed that the population response variable values (hereafter the y-values) are random realizations from some distribution that conditions on known values of auxiliary variables (hereafter the x-values). Classical examples are linear regression and logistic regression models. The problem considered is the prediction of the population total Y of the response variable values based on the sample y-values, the sampling weights for units in the sample and the population values of the x-values. The use of the sample distribution permits the conditioning on all these values and the prediction of Y is equivalent therefore to the prediction of the y-values for units outside the sample. The prediction of the non-sampled y-values is achieved by

approximating the conditional expectation of the y-values given the x-values for units outside the sample as a function of the conditional sample expectation (the expectation under the sample distribution) and the sampling weights. Several parametric and non-parametric procedures are considered for the estimation of the unknown model parameters required for the prediction process. Appropriate estimates for the MSE of the prediction errors are derived. Empirical results comparing the performance of the proposed predictors with other known methods are presented.

Pfeffermann, Danny, Tiller, Richard, and **Zimmerman, Tamara,** (2000), " Accounting for Sampling Error Autocorrelations Towards Signal Extraction from Models with Sampling Error," *Proceedings of the Section on Business and Economics Statistics,* American Statistical Association.

The time series approach to modeling periodic survey data requires information on sampling error autocorrelations (SEA), which, however, are not routinely produced by statistical agencies. Pfeffermann, Feder and Signorelli (1998) have proposed a simple way of estimating SEA from separate panel estimates. This paper examines the properties of this method in more detail and applies it to employment and unemployment series collected by the Current Population Survey (CPS) for 53 States and areas. We present the results of a simulation study designed to assess the performance of the SEA and three alternative variance estimators. One very pleasing result is that the SEA estimators computed from the separate panel estimates have smaller MSE than those computed from the true sampling errors. For variance estimation, the Bootstrap performs best for series of short length and is at least as good as the other methods for long series. To further improve the efficiency of the SEA estimators we borrow information across States using an iterative disjoint clustering algorithm. This algorithm groups States into 2 or 3 clusters. Pooling the estimated State SEA within these clusters results in substantial MSE reductions without introducing significant biases. Finally, we demonstrate how to fit relatively parsimonious ARMA models to the SEA, which is very useful for signal extraction applications.

Polivka, Anne E. and **West, Sandra A**. (1997), "Earnings Data From The Current Population Survey After The Redesign," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

In January 1994, the method for collecting earnings data in the monthly Current Population Survey (CPS) was changed. Prior to January 1994, respondents were forced to report their earnings weekly, and in addition earnings were truncated at $1999 a week. Since January 1994, respondents

have been able to report their earnings in the time frame of their own choosing, and earnings data have not been truncated. In this paper, the effect of the change in the method of collecting earnings data is examined for the comparisons over time of mean and median earnings, for all workers and for various subgroups. In addition, since the micro data released to the public will continue to be truncated, the effect this truncation will be examined and alternative estimators of the mean of the upper earnings distribution will be considered. In earlier papers by West (1986, 1987), a method for computing the mean, which involved fitting a Pareto distribution, was developed, and procedures for computing the mean and median for CPS earnings data were recommended. The appropriateness of these methods are investigated for the earnings data collected after January 1994.

Montaquila, Jill M**.** and **Ponikowski, Chester H.** (1995), "An Evaluation Of Alternative Imputation Methods" *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 281-286.

Several imputation methods have been developed for imputing missing responses. Often it is not clear which imputation method is "best" for a particular application. In choosing an imputation method, one should consider several factors, including the types of estimates that will be generated, the item nonresponse rates, the nature of the missing data, and the availability of auxiliary data that are correlated with the characteristic of interest or with the response propensity. This paper compares the performance of four commonly used imputation methods--nearest neighbor hot-deck, random hot-deck, cell mean, and regression imputation--in imputing missing benefit cost values for the Employment Cost Index (ECI) survey. It is assumed the nonrespondents are missing at random within classes. The advantages and disadvantages of each method are described; analytical and empirical results are presented.

Robertson, Kenneth W., and **Frugoli, Pamela L.,**(1999) " Statistical Issues for the Redesigned Occupational Employment Statistics Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The purpose of this paper is to provide information on changes to the design and uses of the Occupational Employment Statistics (OES) survey. The survey is designed to measure occupational employment and wage rates by industry at the National, State, and Metropolitan Statistical Area (MSA) levels. During the course of the paper, the authors will describe several aspects of the survey. These will include the survey data and its uses, the sample design and estimators, and several statistical issues that

were of interest during the redesign. The following issues are described: (1) combining data across years, (2) allocating the sample, (3) assigning sampling weights, (4) ratio adjusting the estimates to meet multiple estimation goals, (5) updating previous-year wage information, and (6) developing a procedure to calculate medians using grouped wage data from multiple years.

Robertson, Kenneth W., Huff, Larry and **Tou, Albert** (1997), "Developing An Estimator For The Variance Of Mean Wage Rates Computed From Grouped Data In The Occupational Employment Statistics Survey," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 481-486.

The Occupational Employment Statistics (OES) survey is a periodic survey of nonagricultural business establishments conducted by the Bureau of Labor Statistics. In this survey wage data for each occupation are reported by coding the number of employees in each of eleven contiguous, non-overlapping wage intervals. The exact wage rates of individual workers, consequently, are not captured. This paper reports on the use of an optimization approach on an auxiliary data set to derive a set of location parameters for variance estimation. These parameters were used to produce a "parameter set" variance estimator for mean wage rates computed from grouped OES data. We evaluated and compared our "parameter set" variance estimator with a standard "grouped data" variance estimator using exact wage data from another of the Bureau's wage surveys. These exact wage rates were used to simulate the OES grouped data.

Robison, Ed (1999), *"*Sampling and Reporting in Time-Use Surveys," *Proceedings of the Section on Social Statistics,* American Statistical Association.

Many variables potentially affect the way that people use their time. Individuals and families may allocate their time differently based on gender, income, and presence of children, among other factors. In order to report information on time use by these and other variables, a time-use survey sample must be designed that will produce accurate estimates by many of these characteristics. The study of the allocation of time by households or individuals must also examine the tradeoff of time to leisure, paid employment, and other activities. Thus, measures need to be made that examine the use of time during the workweek (or work days) as well as during weekends (or nonwork days). This paper will describe the sample design and reporting issues that were faced when recommending a strategy for a national time-use survey.

Rosen, Richard, Harrell, Louis, **Manning, Christopher** , and **Skuta, Doug** (1999) *"*Data Collection Issues Related to Implementing the Redesigned Current Employment Statistics Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The Current Employment Statistics (CES) program is a monthly survey of over 375,000 business establishments. The survey produces estimates of employment, hours, and earnings for the nation, States, and metropolitan areas. CES data serve as a key measure of the health of the economy. Based on several years of research and the recommendations of a number of expert panels, in mid-1997 BLS launched sample solicitation for a new probability design to replace the existing quota sample. The new sample design is a simple random sample of Unemployment Insurance accounts. Initiating the new sample requires soliciting ongoing reporting from over 240,000 firms. This large-scale undertaking required development of new enrollment protols and procedures. Initiation of the new sample is accomplished primarily via CATI. Ongoing data collection utilizes such automated methods as Touchtone Data Entry (TDE), FAX, Electronic Data Interchange (EDI), and World Wide Web (WWW). In this paper , we discuss the issues involved in implementing data collection. We focus on developing the necessary infrastructure to perform enrollment and data colleciton, development of respondent materials, and management of survey activities across three Data Collection Centers and multiple collection modes. Results to date and implications for other business surveys will be discussed.

Rosen, Richard, O'Connell, **David** and **McKay, Robert** (1997), "Developing An Integrated System For Mixed Mode Data Collection In A Large Monthly Establishment Survey," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 198-203.

The Current Employment Statistics Survey (CES) conducted by the Bureau of Labor Statistics collects employment, payroll, and hours data from a panel of nearly 400,000 businesses each month. These data provide one of the earliest indicators of economic performance, and are closely watched by policy-makers, the business community, and others concerned with the health of the economy. In response to both internal and external pressures, BLS has revamped its data collection procedures. Internal pressures include the need to improve response rates, reduce revisions, and control costs. External pressures come primarily from survey respondents demanding easier less burdensome ways to report. Over the past decade, BLS has transformed the CES from a traditional mail-based survey to utilize an array of automated reporting techniques, including; CATI, touchtone data entry, electronic data interchange, Internet, and FAX. The

goal is to provide the respondent with the fastest most cost-effective reporting method. BLS is now working to integrate each of these methods into a comprehensive data collection system. This paper describes the evolution of automated collection in the CES, reviews the advantages of these new methods, and describes the development of the new integrated survey collection system.

Dillman, D. A., **Sangster, R. L**., Tarnai, J. and Rockwood, T. H. (1996), "Understanding Differences in People's Answers to Telephone and Mail Surveys." *In New Directions in Program Evaluation Series: Current Issues in Survey Research*, Marc T. Braverman and Jana Kay Slater eds. New York: Jossey-Bass, forthcoming. (An earlier version of this paper was presented at the International Conference on Survey Measurement and Process Quality, April 1-4, 1995, Bristol, UK.)

This paper attempts to accomplish three tasks. First, we provide a general model that specifies how mail and telephone data collection methods differ and identify causal mechanisms that seem likely to produce response differences between modes (i.e.. mail and telephone surveys). Second, we identify seven types of mode differences, ranging from social desirability to primacy vs. recency effects, and utilize the model to state hypotheses about each of these possible mode effects. We also review the available literature , and results from unpublished research, to assess the extent of support for those hypotheses. Third, we identify several research issues that need to be addressed in pursuit of the overall goal of understanding both the causes and consequences of mode differences for survey research. The major result of this paper is to provide an explanation for why mode differences may be especially difficult to isolate, and the reasons that considerable inconsistency exists in the available research that has attempted to test for the existence of mode effects.

Dillman , D. A., **Sangster, R. L**., et. al. (1995), "Effects of Category Order on Answers to Mail and Telephone Surveys." *Rural Sociology*, 60(4):674-687.

Past research suggests that mail surveys encourage a primacy effect, which is a tendency to choose the first answers from a list, whereas telephone surveys encourage a recency effect, a tendency to choose the last answers from a list. This paper summarizes results from 82 new experiments conducted by 12 separate surveys in seven states (USA). Only four of 33 mail survey comparison exhibited significant primacy effects, while five of 26 experiments in telephone surveys exhibited recency effects. In addition, only three of 23 cross-method comparisons produced significant primacy/recency effects in the expected manner. The conclusion is that the

prevalence of primacy and recency effects has been over-estimated by past research and that a new theoretical approach which takes into account multiple causation is needed for examining these effects.

Scott, Stuart and **Zadrozny, Peter** (1999) *","* Aggregation & Model-based Methods in Seasonal Adjustment of Labor Force Series *Proceedings of the Section on Business and Economic Statistics,* American Statistical Association.

One of the most closely-watched numbers produced by the U.S. Bureau of Labor Statistics (BLS) is the seasonally adjusted civilian unemployment rate. It is computed from eight employment and four unemployment series. A motivation for examining this issue of aggregation is that four of twelve components of the unemployment rate are for teenagers and four are agricultural employment, series with substantial sampling error. This study revisits work of Estela Dagum (1978).International interest in model-based seasonal adjustment has increased with availability of the TRAMO/SEATS software of Gomez and Maravall (1994). The methodology, which builds on the work of Burman (1980), carries out signal extraction on seasonal ARIMA models. Comparisons will be made between TRAMO/SEATS and X-11 methodology using Census's X-12-ARIMA (Findley et al, 1998).

Schwartz, Lisa and **Paulin, Geoffrey** (2000) "Improving Response Rates to Income Questions: A Comparison of Range Techniques**"** *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Income and other sensitive questions continue to yield high non-response rates on government and non-government surveys. Recent household surveys have implemented the use of follow-up questions with brackets as a partial solution to missing income data. In addition to gaining some information, brackets may provide more accurate imputed values. However, follow-up questions risk increasing respondent burden. The present study directly compares three alternative range techniques: the conventional response card, unfolding brackets, and a new procedure called respondent-generated intervals (RGI) (Press, 1999). Sixty participants were randomly assigned to one of the three range techniques and used that technique to respond to a series of income questions. Participants completed three rating scales that measured their willingness to provide income information, the ease with which they reached an answer and the intrusiveness of the range technique. A structured cognitive interview format was employed to further explore participants' subjective experience of the income interview. During the cognitive interview,

participants sampled and rated each of the other two range techniques under consideration. The results indicated that bracketing questions can be used to reduce nonresponse to income questions. All three techniques generated equally useful income data as indicated by interval width and ability to predict the exact value based on the midpoint of the range. The conventional and unfolding techniques, but not RGI, lost some predictive utility at the upper income levels. The conventional technique was the most preferred whereas the unfolding technique was the least preferred and the most intrusive. Factors underlying participants' preferences and their implications for survey design are discussed.

Pfeffermann, Danny and **Scott, Stuart** (1997), "Variance Measures For X-11 Seasonally Adjusted Estimators; Some New Developments with Application to Labor Force Series," *Proceedings of the Section on Business & Economic Statistics*, American Statistical Association, 211-216

Pfeffermann (1994) has developed a method of estimating variances of X-11 seasonally adjusted series and trend components of time series derived from surveys. With Morry and Wong (1996), he has extended the method to situations where the sampling variances change over time, varying either with the level of the series or due to changes in the sampling design. In this paper the method is further developed, and applied to U.S. Bureau of Labor Statistics employment and unemployment series derived from the Current Population Survey. While not strictly required, information on sampling autocovariances is used to improve performance of the method.

Searson, Michael A. and **Farmer, Tracey** (1997), "Quality Of The BLS Business Establishment List As A Sampling Frame," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 11-20.

The Bureau of Labor Statistics uses the administrative records of the Unemployment Insurance system of the State Employment Securities Agencies as the primary component of its Business Establishment List. This paper will discuss the strengths and weaknesses of this approach and the BLS efforts to enhance and improve the quality of these data for use as a sampling frame. Topics that will be addressed are the use of standardized program software in the states; quality assurance procedures in the industrial coding of establishments; supplemental data collection efforts for multiple establishment firms within a state; and, establishing and maintaining a uniform set of standards for the states are in this paper.

Chaiken, Meredith, Knott, James, Kofner, Aaron and **Shipp, Ken** (1999) "Increased Timeliness of Current Employment Statistics at the Bureau of Labor Statistics," *Proceedings of the Section on Government Statistics,* American Statistical Association.

The Bureau of Labor Statistics (BLS) provides timely and accurate information reflecting the current state of the economy to the public. This monthly data includes the level of employment, aggregated as well as broken out by industrial division, at the national, state, and local level of analysis. These figures reflect the total number of jobs in the economy at a precise moment in time; however, the lag between the collection of data and the publishing date may detract from the usefulness of the information. Recently, the BLS has taken steps to increase its timeliness without sacrificing objectivity or the quality of the estimates. Comparing the preliminary estimates to the final, revised data, we attempt to assess the effects, if any, of an incremental speed-up in production time. Using statewide nonfarm employment data from 1988, 1993, 1996, 1997 and 1998, we study the relative size of the revision of the estimates. Furthermore, we consider the effects at the industry level in the most recent three years. While there is some variation among years and industry, there is not significant evidence to indicate a compromise in quality of the data, while the benefits of a quicker production cycle are quite clear

Shipp, Stephanie, Garner, Thesia, Johnson, David and **Paulin, Geoffrey** (1997), "Developing Poverty Thresholds Using Expenditure Data," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

In 1995, the National Academy of Sciences released a report entitled Measuring Poverty: A New Approach. In this report, suggestions were made to revise the official U.S. poverty thresholds. These thresholds would be based on Consumer Expenditure (CE) Survey data. The purpose of the proposed study is to update the thresholds that were produced in this earlier report with more recent data and to examine the adequacy of the proposed two-parameter equivalence scale used to produce the thresholds. This study will consider the effect that these parameters have on poverty rates, including child poverty rates. Additional thresholds will also be produced; these will account for the flow of services from owned housing and for interarea price differences. In contrast to the earlier study in which Current Population Survey income data were compared to the thresholds to determine poverty counts, we will use income and expenditures from the CE survey.

Shoemaker, Owen and Johnson, William (1999), *"*Estimation of Variance Components for the U.S. CONSUMER PRICE INDEX *," Proceedings of the Section on Government Statistics,* American Statistical Association

> For every new sample for the commodities and services (C&S) component of the U.S. Consumer Price Index (CPI), the Bureau of Labor Statistics attempts to produce a C&S sample design that allocates outlets and quotes in an optimal fashion. This item-outlet optimization C&S sample design requires the estimation of components of variance for the three factors in the design: non-certainty primary sampling units (PSUs), item-strata and outlets. The fourth component of variance is the error term. The total variance of these unit components of variance, divided by their respective number of PSUs, item-strata, outlets and quotes, is then minimized by the optimal number of respective outlets and unique quotes. To produce these components of variance a Random Effects Model was chosen, with the independent random variable for the model an individual price change. Weighted Restricted Maximum Likelihood (REML) estimates were used to calculate the variance components. This paper explicates the methlogy for and the results from these estimates.

Shoemaker, Owen J. (1997), "Performance Comparisons of Laspeyres Indexes with Geometric Mean Indexes in the U.S. CPI," *Proceedings of the Section on Government Statistics*, American Statistical Association, 198-203.

> The official U.S. Consumer Price Index (CPI) is a Laspeyres index. An alternative form of the index known as the geometric mean (GeoMeans) index is scheduled to be published in 1997. In this paper, rental housing data have been studied with regard to rent levels, to the average weights attributed to these levels and to the average rate of rental price change at these levels. These weights are quantity-based. A two-partition model is constructed, with "low" rents compared with "higher" rents. Simulations are run comparing Laspeyres and GeoMeans index results. The differences between the two indexes are plotted against the rate of price change in the "low" rent sector. In the simulations, all the other factors are treated as fixed. A smooth (quadratic) curve results, and the GeoMeans indexes are shown to score consistently higher than the Laspeyres indexes.

Spletzer, James R. (1997), "Longitudinal Establishment Microdata At The Bureau Of Labor Statistics: Development, Uses, And Access," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

> Longitudinal establishment data are potentially very important to the economic policy and research community. Although the cross-sectional

data that are currently produced by the government statistical agencies are invaluable for analysis, the current published statistics lack the ability to describe changes in employment and/or the number of establishments over time. This paper will describe an ongoing research project at the Bureau of Labor Statistics whose objective is to use ES- 202 micro level data to construct a longitudinal database on business establishments. This database will be used to generate high quality, high frequency, timely and historically consistent information regarding job creation and job destruction at the establishment level. Special emphasis will be given to topics of economic interest such as births and deaths of establishments, the life cycle of businesses, the variation of job creation and job destruction over the business cycle, and the share of jobs created by small businesses.

Springer, Glenn, Dorfman, Alan H, Paben, Steven and **Walker, Martha** (1999), *"*Evaluation of Confidence Interval Estimation Methods for the National Compensation Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The National Compensation Survey (NCS) is a Bureau of Labor Statistics (BLS) program which provides data on occupational wages. A theoretical investigation presented by Casady, Dorfman, and Wang (1994 , 1996) suggested that 95% confidence intervals for estimates, when based on the standard normal distribution and standard methods of variance estimation, tend to yield less than the actual 95% coverage. They presented nonstandard methods of constructing confidence intervals, which give more accurate coverage. These intervals tend to be longer than the standard intervals and depend mainly on the use of a t-statistic having degrees of freedom dependent on the available domain data. We modified this method to make it suitable to the multi-stage design of the NCS. Using NCS data, an artificial sampling frame was created and simulated samples were selected. The standard normal confidence intervals were compared to confidence intervals using the t-distribution with weighted degrees of freedom for estimates of means, totals and quantiles. Coverage properties for confidence intervals using the non-standard approach were found to be superior to the standard normal approach.

Zarate, Alvan, Hoy, Easley, **Stamas, George,** and Therriault, Gene D., (2000), "Disclosure Review Board of Federal Agencies: Characteristics, Defining Qualities and Generalizability," *Proceedings of the Section on Government Statistics,* American Statistical Association.

Federal statistical agencies are increasingly concerned about protecting the confidentiality of the data that they collect. In assuring that information

released to the public does not put respondent privacy at risk, they have resorted to a number of mechanisms. Some agencies have relied on ad hoc disclosure risk assessment by one or more persons. Others have designated a single official to review such releases. Still others have asked other agencies to review their occasional releases. During the 1980s, when microdata had become a popular means of data release, the U.S. Census Bureau established a Micro-data Review Panel that worked successfully and became it's present Disclosure Review Board. In recent years, disclosure review boards (DRBs) have been established at several other agencies (NCES, NCHS, BLS). This panel will review the composition and functioning of DRBs at these agencies and discuss the necessary features for any DRB and how they might be generalized to other contexts. In doing so, consideration will be given to factors such as available statistical resources, types and levels of statistical and other skills required, and organizational location.

Stamas, George, **Goldenberg, Karen**, Levin, Kerry and Cantor, David (1997), "Sampling For Employment At New Establishments In A Monthly Business Survey," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 279-284.

The Bureau of Labor Statistics Current Employment Statistics Survey (CES) samples business establishments to produce monthly estimates of employment, hours and earnings by industry for the nation, states and metropolitan areas. Any survey attempting to estimate characteristics of a current population from a dated sampling frame will miss the contribution of new units that did not appear on that frame. For a frame, CES relies on a file of administrative records from state unemployment insurance programs (UI) that are available for sampling at the earliest 6 months after the end of a calendar quarter. The lag in identifying new businesses that come into existence in the first month of the quarter is 9 months. This study considers the possibility of sampling from a supplemental frame of newly established UI accounts that is believed to be more current, including an evaluation of coverage and data quality on the supplemental frame. In our paper, we address the definition of "business birth" and development of a questionnaire to differentiate new employers from others in the sample. We present results from a pilot survey including frame development, estimates of population characteristics, the cost of collecting this information, the timeliness of identifying these employers and other data quality issues.

Stewart, Jay, Goldenberg, Karen, Gomes, Tony, and **Manser, Marilyn,** (2000), " Collecting All-Employee Earnings Data in the Current Employment Statistics

Survey: A Comparison of Two Concepts,"*Proceedings of the Section on Government Statistics,* American Statistical Association.

The Bureau of Labor Statistics' (BLS) Current Employment Statistics (CES) program currently collects earnings and hours data for production and nonsupervisory workers. A review of this program concluded that user needs would be better served if CES collected earnings and hours for all employees using a more comprehensive earnings concept. In a pilot test we compared two approaches to measuring all-employee earnings: (1) gross payroll for the pay period that includes the 12th of the month (payroll includes overtime, but excludes irregular bonuses, commissions, or other nonwage payments) plus a measure of nonwage payments for the entire previous month, and (2) total earnings for the entire previous month (total earnings includes all payments made to employees). Two groups of 500 establishments each were asked to report earnings and the associated hours using concept (1) or concept (2). We found that response rates for gross payroll and total earnings were comparable after controlling for observable differences between the groups. The response rate for nonwage payments was lower than for either gross payroll and total earnings. To assess data quality, we matched reported earnings data to those collected from the Unemployment Insurance records. Again, there was no difference between the two concepts.

Stewart, Jay and **Joyce, Mary** (1999), "*W*hy Do We Need Time-Use Data?," *Proceedings of the Section on Social Statistics,* American Statistical Association

Economists know a lot about individuals' behavior in markets. They have developed theories that explain how individuals respond to change in their economic environment. For example, theory predicts that changes in Social Security that reduce the value of benefits will induce workers to delay retirement. Empirical studies show that prediction is borne out by the data. But there are many areas where economists and other social scientists have well developed theories, but little or no empirical evidence. For example, once people retire, theory predicts that they will engage in more nonmarket production, but there are no data with which to test this hypothesis. In other cases, interesting questions have gone unanswered because of lack of data. For example, how would the distribution of family income be affected if we accounted for the nonmarket production of non-working spouses? In this paper, we discuss some of the uses of time-use data in the context of these and other unanswered questions.

Stewart, Jay and **Frazis, Harley** (1998), "Keying Errors Caused by Unusual Response Categories: Evidence from a Current Population Survey Test," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 131-134.

This paper presents evidence that slight deviations from the usual response categories in CATI questionnaires can cause rates of miskeying that are orders of magnitude higher than found in previous research - about 12 percent in our study. Our data come from a field test of the new education questions in the Current Population Survey, which included a question with two "Yes" categories coded '1' and '2', and a "No" category coded '3' instead of the usual '1' = Yes and '2' = No. This additional "Yes" category apparently confused some interviewers. We corroborated this hypothesis by examining administrative data and by comparing the distribution of responses across interviewers for the problem question and for a similar question with the usual coding using a beta-binomial model. We conclude that survey designers should avoid such unusual response categories.

Stinson, Linda (2000), "'Day of Week' Differences and Implications for Time-Use Research,"*Proceedings of the Section on Social Statistics,* American Statistical Association.

One interesting assessment that can be made with time-use survey data is an analysis of the differences in activity patterns between various days of the week. This study is an attempt to understand the influences of the "day-of-week" variable by comparing daily activity profiles using an index that measures the similarity of activities between pairs of days. The distribution of these differences in time allocation should, in turn, provide contemporary guidance for designing and reporting survey data where "day-of-the-week" is important either as part of the substantive content of the study or as an important descriptive variable.

Moore, Jeff, **Stinson, Linda** and Welniak, Ed. Jr. (2001), "Income Measurement Error in Surveys: A Review," *Journal of Official Statistics*, vol. 16.

Focusing on U.S. Government surveys, we briefly examine two of the more typical quality indicators-- "benchmark" comparisons of survey estimates to independent aggregate estimate, and nonresponse— but focus our attention primarily on response error research which compares individual survey respondents' reports to external measures of truth, often obtained from independent record systems. We also examine recent findings of "cognitive" research into respondents' understanding of the meaning of

income questions, their interpretations of the tasks which income questions present, and attempt to link what we know about income reporting errors to these cognitive processes.

Couper, Mick and **Stinson, Linda** (1999),   "Completion of Self Administered Questionnaires in a Sex Survey," *The Journal of Sex Research*, vol. 36, pp.321-330.

This paper examines completion of four self-administered questionnaires (SAQs) in the 1992 National Health and Social Life Survey, a large national personal-visit survey on sexual behaviors and attitudes in the United States. Based on interviewer reports, only 66% of respondents self-administered all four SAQs, 15% self-administered none of them, and 20% did some of the SAQs themselves. Three factors that may effect the decision to self-administer the SAQs were explored: (a) capacity, (b) motivation, and (c) question sensitivity. The findings suggest it is important to measure who completes the SAQ and to attempt to maximize self-completion of SAQs in surveys.

Moore, Jeff, **Stinson, Linda,** and Welniak, Ed, Jr. (1999),   "Income Reporting in Surveys: Cognitive Issues and Measurement Error." In Monroe Sirken, Douglas Herrmann, Susan Schecter, Norbert Schwarz, Judith Tanur, and Roger Tourangeau (eds.), *Cognitive and Survey Research*. New York: John Wiley & Sons, Inc.

Because income data are germane to a wide array of important policy issues, income questions are almost a constant in government-sponsored surveys. This chapter briefly reviews research on cognitive factors affecting income reporting and research concerning the actual quality of survey reports of income. It's intent is to link knowledge about cognitive factors specific to the reporting of income with knowledge regarding the magnitude and nature of the errors made by respondents in response to income questions.

Moy, Luann and **Stinson, Linda** (1999) "Two Sides of A Single Coin?: Dimensions of Change in Different Settings," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

In March 1998, OMB granted approval to test the wording and format of materials designed to collect aggregate `race' and `Hispanic origin' data according to new standards. The first step in the testing process was to develop three different forms, each approaching the problem of aggregate reporting in a slightly different way. The second step was to test the three forms to determine their strengths and weaknesses and, ultimately, to select

a single form to present to OMB. The first wave of testing was conducted in cognitive labs at the Bureau of Labor Statistics (BLS) and the National Center for Health Statistics (NCHS) by professional staff from 3 different agencies: BLS, NCHS, and the U.S. General Accounting Office (GAO). After assessing the information collected in the lab about the three forms, the next step was to take the form into the field for cognitive testing in the workplace. This paper presents the results of those two waves of testing. It also considers the differences and similarities in the results obtained from cognitive testing in formally different settings and reviews the possible sources of those differences and similarities in findings.

Schecter, Susan, **Stinson**, **Linda** and Moy, Luann (1999), " Developing and Testing Aggregate Reporting Forms for Data on Race and Ethnicity," Paper Presented at the Conference of the Federal Committee on Statistical Methodology.

An interagency team of survey researchers used cognitive methods to develop and test forms that will be used to report aggregated race and ethnicity data. This paper describes the methods and results of the cognitive testing for three unique forms. The paper also addresses data quality issues such as concerns about the incompatibility between the format of the individual source data and the aggregate data. Issues regarding differences between automated and manual data collectors and providers are also addressed.

Stinson, Linda (1999), *" Measuring How People Spend Time,"* *Proceedings of the Section on Social Statistics,* American Statistical Association.

Time-use studies typically have a single focus: the duration of human activities. That is, they ask respondents to report everything they did during a 24-hour period along with the starting and stopping times of those actions. This chronological reporting procedure avoids many of the pitfalls of other survey estimation procedures and is less subject to distortion due to social desirability bias. But there are many methodological considerations to take into account when designing a time-use survey. Decisions concerning reporting procedures and mode of data collection may influence data quality. Likewise, the choice of follow-up probes and the treatment of simultaneous activities can determine the amount of information available for accurate and reliable coding of activities. This paper will describe the methodological decisions with which we were faced when designing a time-use survey and introduce the choices that may be made.

Garner, Thesia, Stinson, Linda and Shipp, Stephanie (1998), "Subjective Assessments of Economic Well-Being: Cognitive Research at the U.S. Bureau of Labor Statistics," *Focus*, vol. 19, pp. 43-46.

Most American adults are confronted daily with the delicate task of balancing their income and expenses, and it is reasonable to expect them to have a rather well-developed budgetary sense. There has been consistent interest in using such perceptions to create measures of economic well-being. Questions, such as, "How much income does it take for you and your family to make ends meet?" are frequently used, but rarely examined for reliability and validity. This paper summarizes BLS cognitive studies focused upon respondents' interpretations of such terms as "minimum income," "necessary expenses," and "making ends meet."

Presser, Stanley and **Stinson, Linda** (1998). "Data Collection Mode and Social Desirability Bias in Self-Reported Religious Attendance," *American Sociological Review*, vol. 63, pp. 137-145.

Compared to conventional interviewer-administered questions about attendance at religious services, self-administered items and time-use items should minimize social desirability pressures. In fact, they each reduce claims of weekly religious attendance by about one-third. This difference in measurement approach does not generally affect associations between attendance and demographic characteristics. It does, however, alter observed trends in religious attendance. In contrast to the almost constant attendance rate recorded by conventional interviewer-administered items, approaches minimizing social desirability bias reveal that weekly attendance has declined continuously over the past three decades. These results demonstrate the role of mode of administration in reducing measurement error.

Stinson, Linda (1997), "Using the Delighted/Terrible Scale to Measure Feelings About Income and Expenses," *Proceedings of the Section on Survey Methods Research,* American Statistical Association, 904-909

The "Delighted/Terrible" (D/T) scale, developed by Andrews and Withey at the University of Michigan, has been used in many quality of life surveys. Based upon their tests, Andrews and Withey suggested that the scale may be effective for measuring respondents' feelings about their income. Using cognitive interviews and focus groups, we looked at ways respondents interpreted the seven response choices in the scale. Our results reveal

some (1) redundancy among response categories, (2) difficulties endorsing the extreme categories, and (3) indications of susceptibility to order effects.

Couper, Mick and **Stinson, Linda** (1999),   "Completion of Self Administered Questionnaires in a Sex Survey," *The Journal of Sex Research*, vol. 36, pp.321-330.

This paper examines completion of four self-administered questionnaires (SAQs) in the 1992 National Health and Social Life Survey, a large national personal-visit survey on sexual behaviors and attitudes in the United States. Based on interviewer reports, only 66% of respondents self-administered all four SAQs, 15% self-administered none of them, and 20% did some of the SAQs themselves. Three factors that may effect the decision to self-administer the SAQs were explored: (a) capacity, (b) motivation, and (c) question sensitivity. The findings suggest it is important to measure who completes the SAQ and to attempt to maximize self-completion of SAQs in surveys.

Steiger, Darby Miller, Mainieri, Tina, and **Stinson, Linda** (1997), "Subjective Assessments of Economic Well-Being: Understanding the Minimum Income Question," *Proceedings of the Section on Survey Methods Research,* American Statistical Association, 899-903.

Several U.S. Surveys, such as the Survey of Income Participation and Programs (SIPP) and the General Social Survey (GSS), as well as the Survey of Consumer Finances from Statistics Canada, have used either a "minimum income" question (MIQ) or a "minimum spend" question (MSQ) to assess people's perceptions of the balance between the money they earn and the money they spend. Based on 48 interviews and 9 focus groups, this paper assesses the methodological and conceptual issues at play when answering the MIQ and MSQ, particularly in the context of the actual survey experience.

Stinson, Linda L. and **Fisher, Sylvia Kay** (1996), **"**Overview Of Data Editing Procedures In Surveys Administered By The Bureau Of Labor Statistics: Procedures And Implications**"** *Proceedings of the first International Computer-Assisted System Information Computing (CASIC) Conference.*

> The purpose of this study is to identify issues associated with cleaning computerized data files, using the example of large-scale data collection activities conducted by the Bureau of Labor Statistics (BLS). This paper will describe data editing activities and processes currently implemented to clean data collected by BLS through CAPI and CATI interviews. Similarities and differences in current data editing procedures are described, as well as common and unique decision rules used by the various

BLS surveys in the conduct of both household and establishment surveys. This study describes features associated with data editing software systems in place at the BLS; identifies data editing issues shared in common among BLS surveys; and documents the types of data editing activities and procedures currently implemented at the BLS, as well as how these procedures address data editing needs. This integrated profile is designed to provide an overview of major data editing activities conducted by the BLS to improve data quality that can enhance and inform data reporting.

Mathiowetz, N. and **Stinson, L.** (1996), **"**The Effect of Length of Recall on the Quality of Survey Data: A Meta-Analytic Approach," Paper presented at the annual American Association for Public Opinion Research Conference, Salt Lake City, Utah. .

Survey research, to a large extent, is dependent upon the retrospective recall ability of respondents. Whether the task is one of recalling the details about a specific episode or estimating the frequency of an activity or event, the respondent must search his or her memory for a response that corresponds to the cue or question. To the extent that an individual is not able to recall the occurrence of an even or details about the event, the quality of the data become questionable. While there has been abundant evidence to support the hypothesis that the quality of retrospective data declines as the length of the recall period increases (e.g., Cannell, et al., 1965), there are also a number of research findings that suggest little to no difference in the quality of retrospective recall as a function of length of recall period (e.g., Mathiowetz and Duncan, 1988). Using meta-analytic techniques, we test the hypothesis that the quality of retrospective reports is a function of the length of recall period. We draw on a body of literature that covers over 50 years and spans a wide range of survey topics.

Presser, S. and **Stinson, L.** (1996), "Estimating the Bias in Survey Reports of Religious Attendance," Paper presented at the annual American Association for Public Opinion Research Conference, Salt Lake City, Utah.

Much of what is know about attendance at religious services comes from survey questions that directly ask respondents about their behavior. The standard Gallup question, for example, asks "Did you yourself happen to attend church or synagogue in the last seven days or not?" and the General Social Survey inquires, "How often do you attend religious services?" Validation studies of other kinds of activities (e.g., voting), as well as one study of church attendance in a small Ohio town, suggest that respondents over report to this kind of item. We present the results of a measurement approach that may overcome the social desirability problem. Our paper

examines religious service attendance reported to a time-use study sponsored by the Environmental Protection Agency. The survey asked respondents to detail everything they did and every place they visited during the preceding 24 hours in order to develop estimates of exposure to various sources of pollutants. This approach should engender little social desirability bias as respondents have no idea the researcher is interested in attendance at religious services (indeed the original research team was not).

Stinson, L., **Fisher, S.** and **Stone, D.** (1996), "An empirical model describing the measurement of educational attainment on the Current Population survey," Paper presented at the American Educational Research Association, New York City, New York.

This paper summarizes the results of a study conducted by the Bureau of Labor Statistics/Census CPS Educational Attainment Redesign Task Force. The Task Force was formed in the summer of 1994 to revise the educational attainment item in the Current Population Survey. The steps taken to revise the item were (1) agreeing on the goals of the revision; (2) developing a prototype questionnaire; (3) cognitive testing of the questions; (4) refining and field testing the questionnaire; and (5) final revisions to the item after analysis of results from the field item and discussions with senior BLS personnel. This paper covers the development of the prototype, results of cognitive testing, results of the field testing, and the final item. .

Swanson, David C, Hauge, Sharon K,. Schmidt, Mary Lynn (1999) "Evaluation of Composite Estimation Methods for Cost Weights in the CPI," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

For each local area and for each item stratum, an estimate of consumer expenditure is needed to weight the market basket of goods and services for which the U.S. Consumer Price Index (CPI) is computed. These expenditure estimates, called *cost weights*, must be computed every time the CPI is revised. Due to small sample sizes in the Consumer Expenditure Survey, high variability is an inherent problem in producing such localized estimates. To alleviate this problem, the U.S. Bureau of Labor Statistics (BLS) employs composite estimation to reduce the mean squared error of the cost weight estimates at the index area/item stratum level. In this paper we summarize the research conducted at the BLS over the past ten years on different methods of composite estimation, and describe the method used in the CPI's 1998 revision.

Tiller, Richard B. (1996), "Time Series Decomposition of Periodic Survey Data with Autocorrelated Errors," *Proceedings of the Section on Business and Economics*, American Statistical Association.

Statistical agencies routinely seasonally adjust a large number of time series based on relatively small samples from periodic surveys with an overlapping design. This type of design generates strong positive autocorrelations in the estimates of key population characteristics, which creates a fundamental identification problem in estimating the unobserved population components of the survey series. Direct application of conventional methods of time series decomposition may fail to adequately separate sampling error (SE) from the population components of interest. From a model-based perspective, the solution is straightforward. SE is treated as an additional unobserved component of the time series, with the special advantage that it is objectively identified by design information. Given this information, a filter is constructed to suppress SE variation along with the seasonal and irregular noise in the population. Most statistical agencies use X-11 and may find full model based estimation difficult to apply to a large number of series. Accordingly, the effects of SE on X-11 estimators are examined and ways of adjusting X-11 are explored. Data from the Current Population Survey are used as examples.

Sirken, Monroe, Tanur, Judith, **Tucker, Clyde N.** and Martin, Elizabeth A. (1997), "Synthesis Of Casm II: Current And Future Directions In Interdisciplinary Research On Cognitive Aspects Of Survey Methods," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 1-10.

The first Advanced Seminar on the Cognitive Aspects of Survey Methodology (CASM I), held in May 1983, was sponsored by the Committee on National Statistics with National Science Foundation funding, CASM I was an outstanding success in achieving its goals of building a bridge between the disciplines of survey research and cognitive psychology, especially survey research. The second CASM Seminar (CASM II), scheduled for June 97, seeks to reinvigorate, expand and accelerate the CASM movement that was launched by CASM I. The theme of CASM I was "building a bridge between the disciplines". The follow-on theme of CASM II is "expanding interdisciplinary survey research". This session will include panelists who served as seminar organizers. They will discuss the theoretical and applied research proposals presented at the seminar and will offer a glimpse of new interdisciplinary survey research projects that are underway or being planned.

Tucker, C. (1996), "Methodological Issues Surrounding the Application of Cognitive Psychology in Survey Research," Paper presented at the International Conference on Social Measurement, University of Essex.

This paper examines issues concerning the design of survey research employing cognitive psychology, the problem of measurement arising in this research, and improvements that could be made to the analytical techniques now being used. The discussion is framed by the principles defining the scientific method -- falsifiability, generalizability, repeatability, and reproducibility.

White, Andrew A., Coffey, Jerry, Carlson, Linda T., Batcher, Mary, **Tucker, Clyde** and Eisinger, Dick (1996), "Customer Satisfaction Surveys: Looking Towards The Future" *Proceedings of the Section on Government Statistics,* American Statistical Association.

In 1993 President Clinton issued Executive Order 12862 which effectively changed the way the Federal Government does business. This Executive directive required Federal agencies to survey customers "to determine the kind and quality of services wanted and their level of satisfaction with existing services." The goal is to bring the Federal Government on par with the best customer service in private sector. Over the past two years, Federal statistical agencies have been involved in many customer survey efforts. Most Federal statistical agencies now have a basic understanding of why customer satisfaction surveys are needed and how to conduct them. But many agencies are beginning to ask for clarification on how to sharpen the focus of customer surveys and link results directly to the achievement of identifiable agency performance goals. This panel of Federal Agency leaders will briefly overview what their agencies have done to develop and utilize customer satisfaction surveys. Panelists will discuss their vision to build and integrate the new customer survey process into existing agency operations, explain how research can be designed to achieve actionable results in support of an agency's mission, and make a recommendations for immediate and future customer survey developments across Federal agency lines.

Valliant, Richard and **Dorfman, Alan H.** (1997), "Stratification On A Size Variable Revisited," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 766-771.

Stratification by size is often used when there is a single auxiliary x and a common model is thought to hold for a full population. One must then decide (a) how to construct the strata, (b) how many strata to form and (c)

how to allocate the sample to the several strata. Recent work by Royall (1992) on weighted balance sheds new light on these old problems. In a class of models that holds for many populations, a stratified weighted balanced sample is optimal in the sense of minimizing the error variance of the best linear unbiased predictor. However, a stratified sample, even with optimal allocation, gains nothing compared to a weighted balanced sample with no stratification. This suggests that stratification on a size variable is a superfluous device in this common situation.

Walker, Ed and Mesenbourg, Tom (1997), "The Census Bureau's Business Register: Quality Issues And Observations," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

The Census Bureau's Standard Statistical Establishment List (SSEL) is a central, multipurpose business register covering the United States' employer business establishments and their parent enterprises. This paper will present brief background information on the SSEL's conceptual and methodological basis, statistical units, data sources, maintenance procedures, and primary uses. Further, it will discuss principal dimensions of business register quality, including coverage, data accuracy, timeliness, relevance, and cost. For each of these dimensions, the paper will highlight critical quality issues, describe what we know from existing quality estimates, assess needs for further evaluation, and suggest opportunities for improvement.

Walker, Martha A. C. and **Bergman, Bruce** (1997), "Estimates Of Year-To-Year Change In Costs Per Hour Worked From The Employer Costs For Employee Compensation Survey," *Proceedings of the Business and Economic Statistics Section*, American Statistical Association.

The Bureau of Labor Statistics' Employer Costs for Employee Compensation (ECEC) survey produces cost level estimates using the data from Employment Cost Index (ECI), which is a leading economic indicator that measures quarterly changes in employer costs for employee compensation. Unlike the ECI which is a Laspeyres, fixed-weight index which eliminates the effects of employment shifts over time among major occupational groups and industries, the ECEC produces information annually on a cost per hour worked of each component of compensation which reflects current employment distributions. This article presents, for the first time, estimates of year-to-year change in the ECEC costs for the components of compensation for private industry workers, by industry division and occupational group, and the corresponding standard errors, together with an analysis of the change estimates. We also briefly review

the sample design and estimation methodology for ECEC and describe how the standard errors were computed for the change estimates using the balanced repeated replication method.

Weber, Wolf, (1999), ″A Method of Microdata Disclosure Limitation based on Noise Infusion and Outlier Substitution," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Recent advances in data processing capabilities are enabling more researchers to analyze large data sets. While increased computing power may result in better research, it also adds to the capability of linking microdata file records to survey respondents, thereby compromising privacy. The Census Bureau's Disclosure Review Board (DRB) is responsible for minimizing the probability of respondent disclosure for Census sponsored or Census collected microdata. Recently the DRB began to enforce a new rule regarding the release of "highly visible" or sensitive variables. The motivation for this study is to suggest an alternate method that will minimize disclosure and will result in more accurate variance and regression_based estimates than those that are obtained by the DRB approved method. Several forms of data disturbance were applied to Consumer Expenditure Survey microdata. This paper describes the method that was submitted to the DRB as a suggested alternative to the current disclosure minimization rules.

Werking, George S. **Jr.** (1997), "Overview Of The CES Redesign Research," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 512-516.

The Current Employment Statistics (CES) program is the nation's largest monthly sample survey with a sample of nearly 400,000 establishments. The monthly data on payroll jobs, hours, and earnings are some of the most closely watched and widely used economic indicators among public and private policy makers. While these data are used quite extensively, concern has always been expressed over the survey's lack of a probability sample design and the timely incorporation of new business births into the sample. In 1993, the Bureau of Labor Statistics (BLS) sponsored a detailed program review of the CES survey by an ASA expert panel. The ASA panel's recommendations, along with results of BLS empirical research on CES error sources lead to a BLS announcement of plans for a full redesign of the CES survey. The CES redesign effort called for a 2-year research phase to be completed in June 1997 to be followed by a 1-year production test of the proposed new methods, procedures, and systems. This paper

discusses the goals of the research program and the Bureau's implementation plans.

Werking, George S., **Clayton, Richard L**. and **Harrell, Louis J. Jr**. (1996), "TDE And Beyond: Feasibility Test Of E-Mail/World Wide Web For Survey Data Collection" *Proceedings of the Section on Survey Research,* American Statistical Association, 768-773.

The Current Employment Statistics (CES) survey collects employment, payroll, and hours data each month from about 400,000 business establishments. The CES is converting totally to automated collection using CATI and TDE. TDE collection is a low cost methodology obtaining very high response rates. TDE has achieved a high degree of respondent acceptance due to its uncomplicated interface. The highly successful TDE methodology can be extended in an "on-line" environment using the World Wide Web (WWW). The CES has developed a prototype E-mail collection system. Paralleling the well-tested TDE methodology, the CES Web approach sends advance notice and NRP messages via E-mail. Respondents connect to our Web site and enter data, explanations, and comments. Data are entered into a forms-based table and e-mailed to BLS. Hypertext links provide additional information to the respondent on definitions and points of contact for assistance. In addition, a respondent package summarizing reporting procedures is provided.

West, Sandra A., Kratzke, Tran and **Grdn, Paul** (1997), "Estimators For Average Hourly Earnings And Average Weekly Hours For The Current Employment Statistics Survey," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 529-534.

In this paper the results of a theoretical and empirical investigation of different estimators for average hourly earnings and average weekly hours for the Bureau of Labor Statistics' Current Employment Statistics (CES) survey are presented. The investigations began in connection with the redesign of the survey. This is a large longitudinal survey of establishments that provides monthly estimates for, among other parameters, total employment, average weekly hours and average hourly earnings for production and non supervisory workers. Ten estimators which, for the most part, are ratios of two random variables are considered. The empirical study was done using national CES data that have responses for relevant characteristics for a period of thirteen months. Estimators were evaluated under both low and high nonresponse rates. In this study, nonrespondents are not missing at random.

Wohlford, John and **Mueller, Charlotte,** (2000), "The Debut of a New Establishment Survey: The Job Openings and Labor Turnover Survey at the Bureau of Labor Statistics ," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The United States has had a strong economic indicator of unused labor supply--the unemployment rate--for decades. However, there has been no parallel measure for unfilled labor demand. In 1999, the Bureau of Labor Statistics (BLS) began developing a new mechanism for producing this long-needed indicator--the Job Openings and Labor Turnover Survey (JOLTS). This new survey will measure a job openings rate and corollary hires and separations rates from a sample of 16,000 business establishments. Late in 2001, monthly estimates of rates and numbers will be published for the nation and for broad North American Industry Classification System (NAICS) sectors. This paper provides a brief history of similar surveys previously conducted by BLS, and discusses the goals of the survey, the concepts used, the sampling procedures, the data collection techniques, the estimation process, and publication plans. It also reviews methodological and data collection research conducted in the development of this new survey. Present throughout the development of concepts, definitions and related methodologies is the overarching goal of providing estimates of a job openings rate directly comparable to the unemployment rate from the Current Population Survey. BLS expects this new indicator to greatly enhance policymakers' understanding of imbalances between the supply and demand of labor.

Woodruff, Steve (1996), "Predicting Births In The Current Employment Statistics Survey," *Proceedings of the Section on Survey Research,* American Statistical Association, 359-362.

The Current Employment Statistics (CES) survey is a monthly survey of about 400,000 business establishments. It is used to estimate total national non-farm employment and other economic statistics. The universe for this survey is known as the ES-202 and is constantly changing due to some establishments going out of business and others starting up. An establishment is defined as an Unemployment Insurance (UI) account on the ES-202 file. The effect of these births (new UI accounts) and deaths (UI accounts that stop reporting) remains an important source of non-sampling error. Karlin's book "A First Course in Stochastic Processes" ,1972, outlines several "Birth and Death Processes" that may describe these changes in the ES-202. Karlin's outline describes birth and death processes without regard to model fitting or parameter estimation. This paper describes maximum likelihood, minimum statistical distance, and nonlinear regression estimators for birth and death process parameters.

Historical ES-202 data up to some time point are used for parameter estimation in a birth and death model and this model is used to predict the number of births that occur for the next few months after this time point. These predictions are compared to births on the ES-202 file and these comparisons are repeated for a variety of time periods, industries, and size classes.

Zadrozny, Peter, (2000), "Modelling Survey-Error Autocorrelations Subject to Time-in-Sample Effects for Model-Based Seasonal Adjustments," *Proceedings of the Section on Business and Economics Statistics,* American Statistical Association.

Survey-error autocorrelations have been ignored in model-based seasonal adjustments of published monthly economic data, causing the adjusted data's trends to be contaminated with survey-error trend components. Using Pfeffermann, Feder, & Signorelli's (JBES, 1998) notion of pseudo survey errors (SE), Tiller (Proc. ASA Bus. & Econ. Stat. Sec., 1996) estimated a conventional, time-invariant, 20 parameter, ARMA(2,17), SE model and used it compute uncontaminated model-based seasonal adjustments. The monthly survey data are generated by a "4-8-4" sampling pattern: respondents are sampled for 4 months, are not sampled for the next 8 months, and then are sampled for 4 more months. The data come as panels of 8 "balanced" responses, one for each of 1 to 8 times in sample. The paper uses the sample design to develop a parsimonious 5-parameter model of each panel unit's responses. The model is periodic, with 8-month periodicity, and implies Tiller's time-invariant ARMA(2,17) process for average panel responses. The model is estimated using a full 8-unit panel data set representing Florida per capita employment. The paper contributes by demonstrating the feasibility of estimating a parsimonious survey-error model based on the 4-8-4 survey design, for the purpose of obtaining uncontaminated model-based seasonal adjustments.

Zadrozny, Peter and Chen, Baoline, (1999), *"*Estimation of Capital and Technology with a Dynmaic Economic Model*," Proceedings of the Section on Business and Economic Statistics,* American Statistical Association.

The paper models firms' decisions on capital and technology as outcomes of a dynamic optimization problem. Current stocks of capital and technology are unobserved and stochastic, determined by last period's net-of-depreciation stocks, current observed investment and research flows, and unobserved disturbances. A quadratic dual-cost function approximates a constant-elasticity production function. Inputs of capital, technology, labor, and materials trade off along convex-to-the-origin isoquants.

Production of saleable output, investment in capital, and research in technology trade off along concave-to-the-origin transformation surfaces. Concavity of the transformation surfaces imposes internal adjustment costs on investment and research, hence, makes capital and technology stocks quasi-fixed. The model is estimated using annual data, from 1947 - 1997, for U.S. total manufacturing. Estimation of capital and technology stocks involves two steps. First, the structural parameters of the model are estimated by maximum likelihood. The missing-data variant of the Kalman filter computes the likelihood function with respect to the observed variables, while allowing capital and technology to be unobserved over the sample period. Second, for given values of estimated parameters, the Kalman smoother obtains sample-period estimates of capital and technology.

Chen, Baoline and **Zadrozny, Peter** (1998), "An Extended Yule-Walker Method for Estimating a Vector Autoregressive Model with Mixed-Frequency Data," *Advances in Econometrics: Messy Data--Missing Observations, Outliers, and Mixed-Frequency Data*, Vol. 13, T.B. Fomby and R.C. Hill (eds.), JAI Press Inc., Greenwich, CT.

Zadrozny (1990) proposed and illustrated a nonlinear Kalman-filtering (KF) method for estimating a vector autoregressive moving-average (VARMA) model with mixed-frequency and partly temporally-aggregated data. The present paper proposes an optimal three-step linear instrumental variable method for estimating a VAR model with mixed-frequency data. The method compensates for missing data arising from mixed frequencies by using restrictions implied by extended Yule-Walker (XYW) equations, beyond the usual minimum YW equations for estimating a VAR model with fully-observed single frequency data. The theory of generalized method of moments is used to derive an asymptotically efficient XYW estimator, determine its asymptotic distribution, and provide asymptotic tests for overidentifying restrictions. The KF method can simultaneously handle missing data, temporal aggregation, measurement errors, reporting delays, and revisions, but performs poorly or not at all on large models with many parameters. The XYW method can handle any pattern of missing data (subject to parameters being identified), but it is not yet clear how it might handle the other mentioned data problems. Having the computational complexity of generalized least squares, the XYW method can handle much larger models. The XYW method is illustrated and compared to the KF method using real and simulated macroeconomic monthly-quarterly data. The large number of available macroeconomic and financial time series at observation frequencies ranging from annual to "tick by tick" offer wide possibilities for applications of the XYW method.

Zimmerman, Tamara, Tiller, Richard, and **Pfeffermann, Danny,** (2000), "Bootstrap Approximation to Prediction MSE for State-Space Models with Estimated Parameters," *Proceedings of the Section on Business and Economics Statistics,* American Statistical Association.

This paper proposes a simple but very general method for approximating the prediction Mean Square Error (MSE) of the state vector predictors estimators in a state-space model, when estimating the unknown model parameters from the observed series. As is well known, substituting the model parameters with the sample estimates in the theoretical MSE formulae that assume known parameter values results in under-estimation of the true MSE. Methods proposed in the literature to deal with this problem are often inadequate, and may not even be operational when fitting elaborate models or when some of the parameters are close to their boundary values. The proposed method consists of generating a large number of realizations from the model fitted to the original observations, re-estimating all the model parameters using the same method as used for the observed values and then modifying the MSE values calculated from the newly generated data sets to obtain the desired MSE approximations. Application of the procedure to a model fitted to a time series of sample estimates of employment ratios from the Current Population Survey (CPS), with eighteen unknown parameters, estimated by a three-step procedure yields very accurate results.